

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 13, 2003, 15:19:10 ; Search time 9.98505 Seconds
(without alignments)
1946.896 Million cell updates/sec

Title: US-09-988-117-3
Perfect score: 5301
Sequence: 1 MAOETSSSRLLGPGCEPAE.....SITSPGSLSPFRKDDPTMP 1002

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 118974 seqs, 19401057 residues

Total number of hits satisfying chosen parameters: 118974

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database : Published Applications AA:*

- 1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
- 2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
- 3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
- 4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
- 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
- 6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
- 7: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
- 8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
- 9: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
- 10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
- 11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
- 12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep:*
- 13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
- 14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	5301	100.0	1002	9 US-09-988-117-3	Sequence 3, Appl1
2	5301	100.0	1002	10 US-09-812-471-3	Sequence 3, Appl1
3	5301	100.0	1002	10 US-09-812-633-3	Sequence 3, Appl1
4	4677.5	88.2	1003	9 US-09-988-117-1	Sequence 1, Appl1
5	4677.5	88.2	1005	10 US-09-812-471-1	Sequence 1, Appl1
6	4677.5	88.2	1005	10 US-09-812-633-1	Sequence 1, Appl1
7	482.5	9.1	330	10 US-09-864-761-42953	Sequence 42953, A
8	438.5	8.3	336	10 US-09-864-761-43157	Sequence 43157, A
9	297.5	5.6	515	10 US-09-864-761-37152	Sequence 37152, A
10	297.5	5.1	577	10 US-09-764-864-883	Sequence 883, App
11	264	5.0	610	9 US-09-989-920-212	Sequence 212, App
12	249.5	4.7	406	10 US-09-764-864-976	Sequence 976, App
13	244.5	4.6	375	10 US-09-864-761-36336	Sequence 36336, A
14	244.5	4.6	513	10 US-09-764-864-1061	Sequence 1061, Ap
15	242.5	4.6	439	10 US-09-864-761-33476	Sequence 33476, A
16	242	4.6	580	10 US-09-925-301-1357	Sequence 1357, Ap
17	242	4.6	2005	10 US-09-735-367B-3	Sequence 3, Appl1
18	242	4.6	2063	10 US-09-735-367B-2	Sequence 2, Appl1
19	241.5	4.6	338	10 US-09-764-864-948	Sequence 948, App

20	239.5	4.5	1325	10 US-09-864-761-35612	Sequence 35612, A
21	238	4.5	342	10 US-09-864-761-36708	Sequence 36708, A
22	237	4.5	551	10 US-09-864-761-37065	Sequence 37065, A
23	236.5	4.5	2783	10 US-09-816-669A-14	Sequence 14, Appl1
24	235.5	4.4	536	10 US-09-864-761-36148	Sequence 36148, A
25	234	4.4	311	10 US-09-864-761-34333	Sequence 34333, A
26	234	4.4	340	10 US-09-864-761-35894	Sequence 35894, A
27	233.5	4.4	983	10 US-09-864-761-38000	Sequence 38000, A
28	232	4.4	361	10 US-09-764-864-893	Sequence 893, App
29	231	4.4	708	10 US-09-764-864-954	Sequence 954, App
30	229	4.3	428	9 US-09-734-329-2	Sequence 2, Appl1
31	227.5	4.3	525	10 US-09-864-761-33653	Sequence 33653, A
32	226	4.3	378	10 US-09-764-864-956	Sequence 956, App
33	226	4.3	492	10 US-09-864-761-42897	Sequence 42897, A
34	226	4.3	516	9 US-09-974-298-118	Sequence 118, App
35	225.5	4.3	383	10 US-09-764-864-1001	Sequence 1001, Ap
36	224.5	4.2	252	10 US-09-816-669A-12	Sequence 12, Appl
37	224.5	4.2	543	9 US-09-736-457-337	Sequence 337, App
38	224.5	4.2	543	9 US-09-902-941-337	Sequence 337, App
39	224.5	4.2	543	9 US-09-849-626-337	Sequence 337, App
40	222	4.2	449	10 US-09-929-315-4	Sequence 4, Appl1
41	222	4.2	449	10 US-09-929-315-4	Sequence 4, Appl1
42	221	4.2	558	10 US-09-864-761-47724	Sequence 47724, A
43	218.5	4.1	284	10 US-09-764-864-932	Sequence 932, App
44	218	4.1	468	10 US-09-864-761-37991	Sequence 37991, A
45	216.5	4.1	510	10 US-09-866-562-56	Sequence 56, Appl
46	216.5	4.1	547	10 US-09-764-864-1065	Sequence 1065, Ap
47	216	4.1	1464	9 US-09-842-777-10	Sequence 10, Appl
48	215.5	4.1	449	10 US-09-929-315-6	Sequence 6, Appl1
49	215.5	4.1	449	10 US-09-929-315-6	Sequence 6, Appl1
50	215	4.1	269	10 US-09-764-864-887	Sequence 887, App
51	214	4.0	333	10 US-09-764-864-947	Sequence 947, App
52	214	4.0	333	10 US-09-764-864-1095	Sequence 1095, Ap
53	214	4.0	448	10 US-09-764-864-909	Sequence 909, App
54	212	4.0	600	10 US-09-764-864-957	Sequence 957, App
55	210	4.0	316	10 US-09-764-864-1036	Sequence 1036, Ap
56	209	3.9	483	10 US-09-764-864-940	Sequence 940, App
57	209	3.9	516	10 US-09-083-290-4	Sequence 4, Appl1
58	208	3.9	432	10 US-09-755-830-39	Sequence 39, Appl1
59	208	3.9	506	10 US-09-083-290-2	Sequence 2, Appl1
60	208	3.9	537	9 US-10-037-667-1	Sequence 1, Appl1
61	208	3.9	1125	9 US-09-974-288-114	Sequence 114, App
62	207	3.9	281	10 US-09-764-864-1337	Sequence 1337, Ap
63	206	3.9	614	10 US-09-864-761-43107	Sequence 43107, A
64	204.5	3.9	488	10 US-09-840-787-17	Sequence 17, Appl
65	203.5	3.8	359	10 US-09-764-864-996	Sequence 996, App
66	203.5	3.8	585	10 US-09-764-864-1063	Sequence 1063, Ap
67	203	3.8	261	10 US-09-764-864-961	Sequence 961, App
68	203	3.8	461	10 US-09-755-850-38	Sequence 38, Appl
69	201.5	3.8	515	9 US-10-037-667-4	Sequence 4, Appl1
70	200	3.8	202	10 US-09-764-864-1337	Sequence 1337, Ap
71	199	3.8	241	10 US-09-764-864-935	Sequence 935, App
72	197.5	3.7	196	9 US-09-942-087A-30	Sequence 30, Appl
73	197.5	3.7	196	10 US-09-779-233-18	Sequence 18, Appl
74	197.5	3.7	196	10 US-09-844-508-29	Sequence 29, Appl
75	197.5	3.7	413	10 US-09-764-864-995	Sequence 995, App
76	197	3.7	826	10 US-09-894-998-47	Sequence 47, Appl
77	195.5	3.7	571	10 US-09-764-864-1028	Sequence 1028, Ap
78	195	3.7	711	10 US-09-976-165-10	Sequence 10, Appl
79	195	3.7	711	10 US-09-828-648-2	Sequence 2, Appl1
80	195	3.7	731	9 US-10-086-464-8	Sequence 8, Appl1
81	194.5	3.7	249	10 US-09-925-300-1179	Sequence 1179, Ap
82	194	3.7	273	10 US-09-764-864-1421	Sequence 1421, Ap
83	193.5	3.7	532	9 US-10-037-667-2	Sequence 2, Appl1
84	193	3.6	277	10 US-09-864-761-36043	Sequence 36043, A
85	193	3.6	361	10 US-09-955-518-21	Sequence 21, Appl
86	193	3.6	481	10 US-09-864-761-37651	Sequence 37651, A
87	192.5	3.6	498	9 US-10-037-667-5	Sequence 5, Appl1
88	192	3.6	488	9 US-09-964-899-35	Sequence 35, Appl
89	192	3.6	723	12 US-10-044-205A-32	Sequence 32, Appl
90	192	3.6	1907	9 US-09-832-292-39	Sequence 39, Appl
91	191.5	3.6	1907	10 US-09-764-864-991	Sequence 991, App
92	191.5	3.6	796	12 US-10-044-205A-31	Sequence 31, Appl

QY 121 GQFLVAAVTGAAGGGGGLILASPKLGATPLPEESTPAPPPPPPPPPGVGSGHLNPLI 180
121 GQFLVAAVTGAAGGGGGLILASPKLGATPLPEESTPAPPPPPPPPPGVGSGHLNPLI 180
QY 181 LEEELRVLOQROIHOMQMEQICROYLLLSLQTVGAPASPSSELPGTGAASSTKPLPLF 240
181 LEEELRVLOQROIHOMQMEQICROYLLLSLQTVGAPASPSSELPGTGAASSTKPLPLF 240
Db 241 SPIPAOTGKTATSSSSSSSGAEPKQAFHLVHPLGSOHFFSVGVGSRSHKPTAPS 300
241 SPIPAOTGKTATSSSSSSSGAEPKQAFHLVHPLGSOHFFSVGVGSRSHKPTAPS 300
QY 301 PALPGSTQOLIASPHLAFPGTTGLLAOCIGAARGLAASPGILKPKNGSGELGYEVI 360
301 PALPGSTQOLIASPHLAFPGTTGLLAOCIGAARGLAASPGILKPKNGSGELGYEVI 360
Db 361 SLEKPGGRHKRCFAKVGSDSALOIHLSRSHGGERPKCNVCGNRFTTRGNLKVHFRH 420
361 SLEKPGGRHKRCFAKVGSDSALOIHLSRSHGGERPKCNVCGNRFTTRGNLKVHFRH 420
QY 421 REKYHVQMNPRVPEHLDYVITSSGLPYGMSVPRPEKAEEDAGTPGGVERKPLVASTTA 480
421 REKYHVQMNPRVPEHLDYVITSSGLPYGMSVPRPEKAEEDAGTPGGVERKPLVASTTA 480
Db 481 LSATESLTLSTGTSTAVAPGLPTFNKFVLMKAVEPKSKADENTPPGSEGAIAGVADSG 540
481 LSATESLTLSTGTSTAVAPGLPTFNKFVLMKAVEPKSKADENTPPGSEGAIAGVADSG 540
QY 541 SATRMQSLKLVSLPFWALLTNHLKSTGSPFPVYLEPLGASPSSETSKLOOLVEKIDRQ 600
541 SATRMQSLKLVSLPFWALLTNHLKSTGSPFPVYLEPLGASPSSETSKLOOLVEKIDRQ 600
Db 541 SATRMQSLKLVSLPFWALLTNHLKSTGSPFPVYLEPLGASPSSETSKLOOLVEKIDRQ 600
QY 601 AVAVASTASGAPTTAPAPSSASGPNQCVICLRVLSCPRALRLHYGONGGERPKKVC 660
601 AVAVASTASGAPTTAPAPSSASGPNQCVICLRVLSCPRALRLHYGONGGERPKKVC 660
Db 601 AVAVASTASGAPTTAPAPSSASGPNQCVICLRVLSCPRALRLHYGONGGERPKKVC 660
QY 661 GRAFSTRGNLRAHFVGHKTSPPAARONSCPTCKKFTNAVYVLOOHVHNLGQIPNGGSA 720
661 GRAFSTRGNLRAHFVGHKTSPPAARONSCPTCKKFTNAVYVLOOHVHNLGQIPNGGSA 720
Db 661 GRAFSTRGNLRAHFVGHKTSPPAARONSCPTCKKFTNAVYVLOOHVHNLGQIPNGGSA 720
QY 721 LSEGGAAOENSSEOSTASPGSPPOSOQPSPEEEMSEEEDEEDVTDSDSLAG 780
721 LSEGGAAOENSSEOSTASPGSPPOSOQPSPEEEMSEEEDEEDVTDSDSLAG 780
Db 721 LSEGGAAOENSSEOSTASPGSPPOSOQPSPEEEMSEEEDEEDVTDSDSLAG 780
QY 781 RGESEGGKAIISVRGDEEVSAGAEVATVAAPTTVKEMDSNEKARPOHTLPPPPPPDN 840
781 RGESEGGKAIISVRGDEEVSAGAEVATVAAPTTVKEMDSNEKARPOHTLPPPPPPDN 840
Db 781 RGESEGGKAIISVRGDEEVSAGAEVATVAAPTTVKEMDSNEKARPOHTLPPPPPPDN 840
QY 841 LDHPQMEQSTDVSGAMEEBAKLEGISSPMAALTOEGEGSTPLVEELNLPKAMKDDG 900
841 LDHPQMEQSTDVSGAMEEBAKLEGISSPMAALTOEGEGSTPLVEELNLPKAMKDDG 900
Db 841 LDHPQMEQSTDVSGAMEEBAKLEGISSPMAALTOEGEGSTPLVEELNLPKAMKDDG 900
QY 901 ESSGRKACEVCGSFPOTALEEHQKTHPKDGLFTCYFCROGFLDRATLKKHMLAHQ 960
901 ESSGRKACEVCGSFPOTALEEHQKTHPKDGLFTCYFCROGFLDRATLKKHMLAHQ 960
Db 901 ESSGRKACEVCGSFPOTALEEHQKTHPKDGLFTCYFCROGFLDRATLKKHMLAHQ 960
QY 961 VPPAPHPGPNIAITLSLVPGCSSSTIPSPGLSPFPKDDPTMP 1002
961 VPPAPHPGPNIAITLSLVPGCSSSTIPSPGLSPFPKDDPTMP 1002
Db 961 VPPAPHPGPNIAITLSLVPGCSSSTIPSPGLSPFPKDDPTMP 1002

RESULT 3
US-09-812-633-3
; Sequence 3, Application US/09812633
; Patent No. US20020147996A1
; GENERAL INFORMATION:
; APPLICANT: Benjamin L.
; APPLICANT: Li, Dawei
; APPLICANT: Mok, Samuel C.
; APPLICANT: Cramer, Daniel W.
; APPLICANT: Ma, Jupo
; TITLE OF INVENTION: Diagnosing and Treating Cancer Cells
; TITLE OF INVENTION: Using Sal2

FILE REFERENCE: 00742/066001
CURRENT APPLICATION NUMBER: US/09/812, 633
CURRENT FILING DATE: 2001-03-19
PRIOR APPLICATION NUMBER: US 60/216,723
PRIOR FILING DATE: 2000-07-07
NUMBER OF SEQ ID NOS: 21
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO: 3
LENGTH: 1002
TYPE: PRF
ORGANISM: Mus musculus
US-09-812-633-3
Query Match 100.0%; Score 5301; DB 10; Length 1002;
Best Local Similarity 100.0%; Pred. No. 4,6e-247;
Matches 1002; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MAQETGSSRLGPGCGEPABERGDASEBHPQVCAKCAQFSDPTFEFLAHQNSCTDPPV 60
1 MAQETGSSRLGPGCGEPABERGDASEBHPQVCAKCAQFSDPTFEFLAHQNSCTDPPV 60
Db 1 MAQETGSSRLGPGCGEPABERGDASEBHPQVCAKCAQFSDPTFEFLAHQNSCTDPPV 60
QY 61 MVIIGGDNPSNSASAPRDEGHSRSQVMDTEHNSNPDSGSSGPPDPTWGPERRGEES 120
61 MVIIGGDNPSNSASAPRDEGHSRSQVMDTEHNSNPDSGSSGPPDPTWGPERRGEES 120
Db 61 MVIIGGDNPSNSASAPRDEGHSRSQVMDTEHNSNPDSGSSGPPDPTWGPERRGEES 120
QY 121 GQFLVAAVTGAAGGGGGLILASPKLGATPLPEESTPAPPPPPPPPPGVGSGHLNPLI 180
121 GQFLVAAVTGAAGGGGGLILASPKLGATPLPEESTPAPPPPPPPPPGVGSGHLNPLI 180
Db 121 GQFLVAAVTGAAGGGGGLILASPKLGATPLPEESTPAPPPPPPPPPGVGSGHLNPLI 180
QY 181 LEEELRVLOQROIHOMQMEQICROYLLLSLQTVGAPASPSSELPGTGAASSTKPLPLF 240
181 LEEELRVLOQROIHOMQMEQICROYLLLSLQTVGAPASPSSELPGTGAASSTKPLPLF 240
Db 181 LEEELRVLOQROIHOMQMEQICROYLLLSLQTVGAPASPSSELPGTGAASSTKPLPLF 240
QY 241 SPIPAOTGKTATSSSSSSSGAEPKQAFHLVHPLGSOHFFSVGVGSRSHKPTAPS 300
241 SPIPAOTGKTATSSSSSSSGAEPKQAFHLVHPLGSOHFFSVGVGSRSHKPTAPS 300
Db 241 SPIPAOTGKTATSSSSSSSGAEPKQAFHLVHPLGSOHFFSVGVGSRSHKPTAPS 300
QY 301 PALPGSTQOLIASPHLAFPGTTGLLAOCIGAARGLAASPGILKPKNGSGELGYEVI 360
301 PALPGSTQOLIASPHLAFPGTTGLLAOCIGAARGLAASPGILKPKNGSGELGYEVI 360
Db 301 PALPGSTQOLIASPHLAFPGTTGLLAOCIGAARGLAASPGILKPKNGSGELGYEVI 360
QY 361 SLEKPGGRHKRCFAKVGSDSALOIHLSRSHGGERPKCNVCGNRFTTRGNLKVHFRH 420
361 SLEKPGGRHKRCFAKVGSDSALOIHLSRSHGGERPKCNVCGNRFTTRGNLKVHFRH 420
Db 361 SLEKPGGRHKRCFAKVGSDSALOIHLSRSHGGERPKCNVCGNRFTTRGNLKVHFRH 420
QY 421 REKYHVQMNPRVPEHLDYVITSSGLPYGMSVPRPEKAEEDAGTPGGVERKPLVASTTA 480
421 REKYHVQMNPRVPEHLDYVITSSGLPYGMSVPRPEKAEEDAGTPGGVERKPLVASTTA 480
Db 421 REKYHVQMNPRVPEHLDYVITSSGLPYGMSVPRPEKAEEDAGTPGGVERKPLVASTTA 480
QY 481 LSATESLTLSTGTSTAVAPGLPTFNKFVLMKAVEPKSKADENTPPGSEGAIAGVADSG 540
481 LSATESLTLSTGTSTAVAPGLPTFNKFVLMKAVEPKSKADENTPPGSEGAIAGVADSG 540
Db 481 LSATESLTLSTGTSTAVAPGLPTFNKFVLMKAVEPKSKADENTPPGSEGAIAGVADSG 540
QY 541 SATRMQSLKLVSLPFWALLTNHLKSTGSPFPVYLEPLGASPSSETSKLOOLVEKIDRQ 600
541 SATRMQSLKLVSLPFWALLTNHLKSTGSPFPVYLEPLGASPSSETSKLOOLVEKIDRQ 600
Db 541 SATRMQSLKLVSLPFWALLTNHLKSTGSPFPVYLEPLGASPSSETSKLOOLVEKIDRQ 600
QY 601 AVAVASTASGAPTTAPAPSSASGPNQCVICLRVLSCPRALRLHYGONGGERPKKVC 660
601 AVAVASTASGAPTTAPAPSSASGPNQCVICLRVLSCPRALRLHYGONGGERPKKVC 660
Db 601 AVAVASTASGAPTTAPAPSSASGPNQCVICLRVLSCPRALRLHYGONGGERPKKVC 660
QY 661 GRAFSTRGNLRAHFVGHKTSPPAARONSCPTCKKFTNAVYVLOOHVHNLGQIPNGGSA 720
661 GRAFSTRGNLRAHFVGHKTSPPAARONSCPTCKKFTNAVYVLOOHVHNLGQIPNGGSA 720
Db 661 GRAFSTRGNLRAHFVGHKTSPPAARONSCPTCKKFTNAVYVLOOHVHNLGQIPNGGSA 720
QY 721 LSEGGAAOENSSEOSTASPGSPPOSOQPSPEEEMSEEEDEEDVTDSDSLAG 780
721 LSEGGAAOENSSEOSTASPGSPPOSOQPSPEEEMSEEEDEEDVTDSDSLAG 780
Db 721 LSEGGAAOENSSEOSTASPGSPPOSOQPSPEEEMSEEEDEEDVTDSDSLAG 780
QY 781 RGESEGGKAIISVRGDEEVSAGAEVATVAAPTTVKEMDSNEKARPOHTLPPPPPPDN 840
781 RGESEGGKAIISVRGDEEVSAGAEVATVAAPTTVKEMDSNEKARPOHTLPPPPPPDN 840
Db 781 RGESEGGKAIISVRGDEEVSAGAEVATVAAPTTVKEMDSNEKARPOHTLPPPPPPDN 840

QY	LDPPDMEGTSIVNSGAMEEBEAKLEGISPPAAALTOEEBEGSTPLVEELNEDEAMKDDG	900
84.1	LDPPDMEGTSIVNSGAMEEBEAKLEGISPPAAALTOEEBEGSTPLVEELNEDEAMKDDG	900
Db	LDHPDMEGTSIVNSGAMEEBEAKLEGISPPAAALTOEEBEGSTPLVEELNEDEAMKDDG	900
84.1	LDHPDMEGTSIVNSGAMEEBEAKLEGISPPAAALTOEEBEGSTPLVEELNEDEAMKDDG	900
QY	ESSGRKACEVCGGSPFQTLALBEHQKTHPKDGPFLTYCFYCRGGELIDRATLKHHMLAHHQ	960
90.1	ESSGRKACEVCGGSPFQTLALBEHQKTHPKDGPFLTYCFYCRGGELIDRATLKHHMLAHHQ	960
Db	ESSGRKACEVCGGSPFQTLALBEHQKTHPKDGPFLTYCFYCRGGELIDRATLKHHMLAHHQ	960
90.1	ESSGRKACEVCGGSPFQTLALBEHQKTHPKDGPFLTYCFYCRGGELIDRATLKHHMLAHHQ	960
QY	VPPFAPHGPNITLISLVPGCCSSSIPSPGSLSPFRKDDPTMR	1002
96.1	VPPFAPHGPNITLISLVPGCCSSSIPSPGSLSPFRKDDPTMR	1002
Db	VPPFAPHGPNITLISLVPGCCSSSIPSPGSLSPFRKDDPTMR	1002
96.1	VPPFAPHGPNITLISLVPGCCSSSIPSPGSLSPFRKDDPTMR	1002

```

? RESULT 4
? US-09-988-117-1
? Sequence 1, Application US/09988117
? Patent No. US20020156039A1
? GENERAL INFORMATION:
? APPLICANT: Benjamin, Thomas L.
? APPLICANT: Li, Dawei
? APPLICANT: Mok, Samuel C.
? APPLICANT: Cramer, Daniel W.
? APPLICANT: Ma, Yupo
? TITLE OF INVENTION: Diagnosing and Treating Cancer Cells
? TITLE OF INVENTION: Using Sal2
? FILE REFERENCE: 00742/066002
? CURRENT APPLICATION NUMBER: US/09/988,117
? CURRENT FILING DATE: 2001-11-16
? PRIOR APPLICATION NUMBER: US 09/812,633
? PRIOR FILING DATE: 2001-03-19
? PRIOR APPLICATION NUMBER: US 60/216,723
? PRIOR FILING DATE: 2000-07-07
? NUMBER OF SEQ ID NOS: 21
? SOFTWARE: FastSeq for Windows Version 4.0
? SEQ ID NO 1
? LENGTH: 1005
? TYPE: PRT
? ORGANISM: Homo Sapiens
? US-09-988-117-1

```

Query Match	Similarity	88.2%	Score 4677.5	DB 9	Length 1005
Best Local	Similarity	89.0%	Pred. No. 3.3e-217		
Matches	894	Conservative	25	Mismatches	83
				Indels	3
				Gaps	3
QY	1	MAQETGSSRLGPGCGEPAERNGDASEEHNDPVACAKCAQESDPTEFLAHONSCCTDPV	60		
	1				
DB	1	MAHESERSRLGVPRGEPAELGGASEEDHDPVCAKCAQFDPTEFLAHONACSTDPRV	60		
QY	61	MYIIGGQENPENSASSAPPEGHSRSQVMYDTEHNSNPDSGSGSPDDTPWGERGEESS	120		
	61				
DB	61	MYIIGGQENPNNSASSAPPEGCHNNPQVMDTEHNSNPDSGGSVPTDTPWGERGEESS	120		
QY	121	GQPLVAALGTAAAGGGGGLILASPKIGATPLPESTPAPAPPPPPPPGVSGHNTPLI	180		
	121				
DB	121	GHLVLAATGTAAAGGGGGLILASPKIGATPLPESTPAPAPPPPPPPGVSGHNTPLI	180		
QY	181	LEELRVLQOROIHMOMTEQICROYVLLIGSLQIVGAPASBELPDTGAASSTKPLPLE	240		
	181				
DB	181	LEELRVLQOROIHMOMTEQICROYVLLIGSLQIVGAPASBELPDTGAASSTKPLPLE	240		
QY	241	SPIKPAQTKTTPA-SSSSSSSSSGAEPKQAFHLYHPLGSGHPRPSVGVGSHKPTPAP	299		
	241				
DB	241	SPIKPVGTSKTLASSSSSSSSSGAETPKQAFHLYHPLGSGHPRPSAGVGVSHKPTPAP	300		
QY	300	SPALPGSTDOLILASPHLAFPTGTGILAAOCGAAGLMAAASPGILKRNKGEGELGYEV	359		
	301				
DB	301	SPALPGSTDOLILASPHLAFPTGTGILAAOCGAAGLMAAASPGILKRNKGEGELSTGEV	360		
QY	360	TSLSLEKPGGRHKRCPCAKVFGSDSALQIHLNSHTGERPYKCNCVGNRPTTTRGNLKVPHR	419		
	361				
DB	361	MGPLKEPGGRHKRCPCAKVFGSDSALQIHLNSHTGERPYKCNCVGNRPTTTRGNLKVPHR	420		
QY	420	HREKYPHNOAMPHEVPHILDVITSSGPGYCMASVPEKAEAEAGTGGGVGVERKPLVASTT	479		

Dd	421	HRKPIPHQMPHVPPEHLDVYITSSGLPIGMSVPPKAEBEATPEGGYERKPLVAST	480
Oy	480	ALSATESITLLSTSTSTAVAPAGLPTFNKFLYMKAVEKSKADENTPEGSGASATAGVDS	539
Dd	481	ALSATESITLLSTSTAGTATAHGLPAFNKFLYMKAVEKKNKADENTPEGSGASATAGVDS	540
Oy	540	GSATPMQSLKVTSLSPALLTNLKTSGSPFFVYLEPLGASPSFSTKQQLYEKIDRQ	599
Dd	541	STATLMOJSLKMTSLSPALLTNHFKTSGSPFLPCARALGASPSFSTKQQLYEKIDRQ	600
Oy	600	GAVAAVASTASGAPTTSAVAPSSA-SGPNOCVICTRLVSCPRALRLHYGOHGERPPCK	658
Dd	601	GAVAVTSAASASAPTTSAVAPSSASGPNOCVICTRLVSCPRALRLHYGOHGERPPCK	660
Oy	659	VCGRAFSTRGNLRHAFVGHKTSPPARAONSCPIQCKFTNAVITLQOHVNMHLCQJIPNG	718
Dd	661	VCGRAFSTRGNLRHAFVGHKTSPPARAONSCPIQCKFTNAVITLQOHVNMHLCQJIPNG	720
Oy	719	SALSPGGGAAGENSSEOSTASPGSPFOPOQSOQSPREEMS-EEDEEDEEEDVDHDS	777
Dd	721	TALPBGGGGAAGENSSEOSTASPGSPFOQSOQSPREELSSEEEDEEEDVDHDS	780
Oy	778	LAGRSESGGKKAISVYKDSSEVYSGAEEVAITSVAAPITVYKEMDSNEKAOHTLP PPP	837
Dd	781	LAGRSESGGKKAISVYKDSSEVYSGAEEVOTVAAMTACKEMDSNEKTOQSSLP PPP	840
Oy	838	PDNDLHPHMGOGTSDVYSGAMEEBAKLEGISSPMAALTORGEGSTPLVLELNULEPMKK	897
Dd	841	PDSDLOPDPMGOGSSGVYGGKEEGKREBRSSSPASALTPEGEAITSVLYBELSLQEMAKR	900
Oy	898	DPESSSGKKAEEVCGSPFTTALKEBHOQTHPKGAPLFTVCFCROGFLDRATLKKHMLA	957
Dd	901	EPGESSSKKAEEVCGAFPSGALAEHOKTHPKGAPLFTVCFCROGFLDRATLKKHMLA	960
Oy	958	HHOYPPAPHGPONITLSTLVGCGSSSTPSEGLSPPFRKDDPTMP 1002	
Dd	961	HHOYPPAPHGPONIAALSTLVGCGSPSTSTTSGLSPPFRKDDPTIP 1005	

```

RESULT 5
US-09-812-471-1
? Sequence 1, Application US/09812471
? Patent No. US20020018765A1
? GENERAL INFORMATION:
? APPLICANT: Benjamin, Thomas L.
? TITLE OF INVENTION: Diagnosing and Treating Cancer Cells
? TITLE OF INVENTION: Using Mutant Viruses
? FILE REFERENCE: 00742/062002
? CURRENT APPLICATION NUMBER: US/09/812.471
? CURRENT FILING DATE: 2001-03-19
? PRIOR APPLICATION NUMBER: US 60/216,723
? PRIOR FILING DATE: 2000-07-07
? NUMBER OF SEQ ID NOS: 21
? SOFTWARE: FastSeq for Windows Version 4.0
? SEQ ID NO 1
? LENGTH: 1005
? TYPE: PRT
? ORGANISM: Homo Sapiens
US-09-812-471-1

```

Query Match	88.2%	Score 4677.5	DB 10	length 1005
Best Local Similarity	89.0%	Pred. No 3.3e-217		
Matches	894	Conservative 25	Mismatches 83	Indels 3; Gaps 3;
QY	1	MAOETGSSRLGGPCGEPAERGGDASEHHNPVCAKCAQNSDPLEFLAHONSCCTDPV	60	
	1			
Db	1	MAHESRSRLGVPGEPAELGGDASEEDHPVCAKCAQNPDPLEFLAHONACSTDPPV	60	
QY	61	MYLIGQENPNSNSASAPREBCHRSQVMDTEHNPDPDGGSGGPPDPTWCPERGEESS	120	
	61			
Db	61	MYLIGQENPNSNSASAPREBCHHNPVMDTEHNPDPDGGSGSVTDPDTPWCPERGEESS	120	

QY	121	GOFLVATGTAAGGGCGLLASP	KLCAATPLPEBSTAPAPPPPPPPPPPPGCGSHLNP	180
QY	121	GHLVATGTAAGGGCGLLASP	KLCAATPLPEBSTAPAPPPPPPPPPPPGCGSHLNP	180
QY	181	LEELRVLOOROIHOMOMEOIC	ROYVLLSLGOTVGAAPASELPJGCAASSTRPLPLF	240
QY	181	LEELRVLOOROIHOMOMEOIC	ROYVLLSLGOTVGAAPASELPJGCAASSTRPLPLF	240
QY	241	SPKPAOTGKTTA-SSSSSSSS	SGAEPKQAFPHLYHPLGSQHPFSVGVGRSHKPTAP	299
QY	241	SPKPAOTGKTTA-SSSSSSSS	SGAEPKQAFPHLYHPLGSQHPFSVGVGRSHKPTAP	299
QY	300	SPAPGSTDOLIASHPLAPGT	TGLAACCLGAARGLAASPELPLKNGSGELGYEV	359
QY	301	SPALPGSTDOLIASHPLAP	GTGLAACCLGAARGLAASPELPLKNGSGELGYEV	360
QY	360	ISSLEKPGGRHKRCACAKV	FGSDALQILHRSHTGERPYKCMVGNRPTTGNLKVHEHR	419
QY	361	MGPLEKPGGRHKRCACAKV	FGSDALQILHRSHTGERPYKCMVGNRPTTGNLKVHEHR	420
QY	420	HREKTPHVOQMNHPREHLD	YVITSSGLPRGMSVPRPKAEEEAATPGGCVKRPVAST	479
QY	421	HREKTPHVOQMNHPREHLD	YVITSSGLPRGMSVPRPKAEEEAATPGGCVKRPVAST	480
QY	480	ALSATESLTLSTGSTVAAP	GLPTENKFLYLMKAVEBKSKADEMTPPSESALAGVADS	539
QY	481	ALSATESLTLSTGSTVAAP	GLPTENKFLYLMKAVEBKSKADEMTPPSESALAGVADS	540
QY	540	GSATMOLSKLVTSLEP	SMALLTNHLKSTGSPFPYVLEPLGASFSETSKLQOLVAKIDRO	599
QY	541	STATMOLSKLVTSLEP	SMALLTNHLKSTGSPFPYVLEPLGASFSETSKLQOLVAKIDRO	600
QY	600	GAVAVASTASGAPPTTAP	APASSA-SGPNOCVICLRVLCRPAALRLAHGONGGERPFCK	658
QY	601	GAVAVASTASGAPPTTAP	APASSA-SGPNOCVICLRVLCRPAALRLAHGONGGERPFCK	660
QY	659	VCGRAFTSGNLRAHFVG	KTSPARAONSCPTOCKFTNAVTLQOAHYRMLHGGQIPNGG	718
QY	661	VCGRAFTSGNLRAHFVG	KTSPARAONSCPTOCKFTNAVTLQOAHYRMLHGGQIPNGG	720
QY	719	SALSEGGGAOENSSBES	TSASGSPSPFOPOGQSPSEEMS-EDEEDEDDEEDYDDEDS	777
QY	721	TALPREGGGAOENSSBES	TSASGSPSPFOPOGQSPSEELS-EBEEDDEDEEDYDDEDS	780
QY	778	LAGRGSESGEKEAISVR	KDSEEVSGAEAEVATSVAAPTTVKEMDSNEKAPOHTLPPPPP	837
QY	781	LAGRGSESGEKEAISVR	KDSEEVSGAEAEVATSVAAAGTMDNEKTYTOOSLPPPPP	840
QY	838	PNNLDHPORNEOCTSDV	SGAMEEBAKLGTSPPMAALQOEGGCTSTPLVEELNLEPAMKK	897
QY	841	PNNLDHPORNEOCTSDV	SGAMEEBAKLGTSPPMAALQOEGGCTSTPLVEELNLEPAMKK	900
QY	898	DGSESGSRKACEVCGSG	FPQTALAEHOKTTPKKGPLUTCVCROGFLDRLATLKKHMLLA	957
QY	901	EPRESSSRKACEVCGSG	FPQGAALAEHOKTTPKKEBPLUTCVCROGFLERATLKKHMLLA	960
QY	958	HHOVPAPAPHPONIAL	TLVLPGCCSSSTPSPGLSPFPKDDPTMP	1002
QY	961	HHOVPAPAPHPONIAL	TLVLPGCCSPITSTGLSPFPKDDPTIP	1005
RESULT 6				
US-09-812-633-1				
Sequence 1, Application US/09812633				
Patent No. US20020147996A1				
GENERAL INFORMATION:				
APPLICANT: Benjamin, Thomas L.				
APPLICANT: Li, Dawei				
APPLICANT: Mok, Samuel C.				
APPLICANT: Cramer, Daniel W.				
APPLICANT: Ma, Yupo				
TITLE OF INVENTION: Diagnosing and Treating Cancer Cells				
TITLE OF INVENTION: Using Sal2				

```

; FILE REFERENCE: 00742/066001
;
; CURRENT APPLICATION NUMBER: US/09/812,633
; CURRENT FILING DATE: 2001-03-19
; PRIOR APPLICATION NUMBER: US 60/216,723
; PRIOR FILING DATE: 2000-07-07
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO. 1
; LENGTH: 1005
;
; TYPE: PRT
; ORGANISM: Homo Sapiens
;
US-09-812-633-1

```

Query Match	88.2%	Score 4677.5	DB 10	Length 1005
Best Local Similarity	89.0%	Pred. No. 3,3e-217		
Matches 894	Conservative 25	Mismatches 83	Indels 3	Gaps 3
QY 1	MAOETSSSRKLGPCGPEAPERGDSAEENHPQVCAKCAQCPDSDETFEIAHONSCCTDEPV	60		
DB 1	MAHESERRSKLGPAPGAPALGDSAEEDHPQVCAKCAQCTDTEFLAHQNACTDEPV	60		
QY 61	MVIIGQENPNSSASAPRPEGHRSQVMDTEHSNPPDSSGCPDPTWGPBERGESS	120		
DB 61	MVIIGQENPNSSASAPRPEGHRSQVMDTEHSNPPDSSGCPDPTWGPBERGESS	120		
QY 121	GQPLVATGTAAGGGGLILASPKLGATPLPESTPAPRPPPPPPPGVSGHINPLI	180		
DB 121	GHPVATGTAAGGGGLILASPKLGATPLPESTPAPRPPPPPPPGVSGHINPLI	180		
QY 181	LEELRYLQQRQIHOMQMTQICRQVLLGSLQTVGAPASSELPTGTAASSTKPLPLF	240		
DB 181	LEELRYLQQRQIHOMQMTQICRQVLLGSLQTVGAPASSELPTGTAASSTKPLPLF	240		
QY 241	SPIKPATGTTLA-SSSSSSSSGAEPKQAFPHLYHPLGSHPPSVGVSCHKPTAP	299		
DB 241	SPIKPATGTTLA-SSSSSSSSGAEPKQAFPHLYHPLGSHPPSVGVSCHKPTAP	300		
QY 300	SPALPGSTDLILASPHLAFPGTTGILAAQCLGAARGLBAASPGLLKPKNSGELGYEV	359		
DB 301	SPALPGSTDLILASPHLAFPGTTGILAAQCLGAARGLBAASPGLLKPKNSGELGYEV	360		
QY 360	ISLEKRGGRKRCFKPCKVFGSDSALOHLHSHNGERYKCAVCNCRFTTRCNLKKVHFR	419		
DB 361	MGPLEKRGGRKRCFKCAVFGSDSALOHLHSHNGERYKCAVCNCRFTTRCNLKKVHFR	420		
QY 420	HREKYPHQMPHPVPEHLIDVITSSGLPYGMSVPERKAEEBAGTPGGVERKPLVASTT	479		
DB 421	HREKYPHQMPHPVPEHLIDVITSSGLPYGMSVPERKAEEBAGTPGGVERKPLVASTT	480		
QY 480	ALSATTESTLLSTGTSTAVAPGLPTFNKFFVLMKAVEPKSAKADENTPPGSEGSALACVADS	539		
DB 481	ALSATTESTLLSTGTSTAVAPGLPTFNKFFVLMKAVEPKSAKADENTPPGSEGSALACVADS	540		
QY 540	GSARPMOISKLVTSILPSAALLTNHLKSGSPFPFVLEPLGASPSSETSKLOQVEKIDRQ	599		
DB 541	STALIMQISKLVTSILPSAALLTNHLKSGSPFPFVLEPLGASPSSETSKLOQVEKIDRQ	600		
QY 600	GAVAVASTASGAPPTSADAPSSA-SGPNOVCICLVILSCPRALRLHYGONGGERPPCK	658		
DB 601	GAVAVASTASGAPPTSADAPSSA-SGPNOVCICLVILSCPRALRLHYGONGGERPPCK	660		
QY 659	VCGRAFSTRGNTRAFVGHKTSPPAARAQNSCPTQCKFFNAYTLQOHVHRHNLGQIPNGG	718		
DB 661	VCGRAFSTRGNTRAFVGHKTSPPAARAQNSCPTQCKFFNAYTLQOHVHRHNLGQIPNGG	720		
QY 719	SALSGGCAAOENSSSQSTASGPGSPQPOSOOPSPDEEMS-BEEDDEBEEDVTDSDS	777		
DB 721	TALPBGGAQONGSQSTVSGAGSPQPOSOOPSPDEEMS-BEEDDEBEEDVTDSDS	780		
QY 778	LARGSGSGEKAIVRGDSEVSGAEEVAIVSVAAPTYVKEMDSNEKAPQHLLRPPPP	837		
DB 781	LARGSGSGEKAIVRGDSEVSGAEEVGVYAAATAGKEMDSNEKTTQOSSLRPPPP	840		

[illegible]

```

OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.2
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.4
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.4
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 0.96
OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 0.94
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.97
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.5
OTHER INFORMATION: EST. HUMAN HIT: A0131453.1, EVALU1 1.00e-105
US-09-864-761-42953

Query Match          9.1%; Score 482.5; DB 10; Length 330;
Best Local Similarity 35.7%; Pred. No. 8,8e-17;
Matches 131; Conservative 37; Mismatches 98; Indels 101; Gaps 13;

OY 531 SATAGVADSSATRMOLSKLVTSLPMSALLTNHLKSTGSFPFPYVLEPFGASPSSTKLQ 550
DB 4 SPRAADGPPASATTF-----TNPLLPIMSEQFA--KFPFGGLD--SAQASSTKLQ 52
OY 591 QLVKEDIDRGAVNAVASTVASCAPFTTSAPAPSSASGPNOCVILRLVSCPRALRLHYGOHG 650
DB 53 QLVENIDKK-----ATDPNCCIHRVLVSCGSA LKMHYRTH 89
OY 651 GERPEFCVKVCGRAPFTNGNLRAHFVGHKTSPPAARONSCPIOCKFTNATVLOQHYRML 710
DB 90 GERPEFCCKICGRAFTTNGNLKTHYSVHRAMPPLRVQHSCTPCKKFTNATVLOQHIRMM 149
OY 711 GGGIPN-----GGSALSTEGGAOEN-----SSRQSTASGSPFQ-POSQOPSP 755
DB 150 GGOIPNTPPVYDSYSESNESDTGSFDEKKNFDDLDNDSDENMECPGSIPTDTPSADASQD 209
OY 756 E-----EMSE-----EEDEDEEEDVTDSDSLARG-- 782
DB 210 SLSSSPLPLEMSSIALENOMKMINAGLAEOQLASLKSVEGNSIGDVTLTNPSSVSGDM 269
OY 783 -SRSGEKAATSVAGDSREV---SGAEEFVATSVAAFTYKEDNSEKARQHTLPP- 834
DB 270 ESGSASPALSTESTSSQALSPSNSSTQDFHKS-----PSTEEKQRAVPSEFANG 319
OY 835 -PPPPDN 840
DB 320 LSPTPVN 326

RESULT 8
US-09-864-761-43157
/ Sequence 43157, Application US/09864761
/ Patient No. US20020048763A1
/ GENERAL INFORMATION:
/ APPLICANT: Penn, Sharron G.
/ APPLICANT: Rank, David R.
/ APPLICANT: Hanzel, David K.
/ APPLICANT: Chen, Wenshang
/ TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
/ FILE REFERENCE: Aeomica-X-1
/ CURRENT APPLICATION NUMBER: US/09/864,761
/ CURRENT FILING DATE: 2001-05-23
/ PRIOR APPLICATION NUMBER: US 60/180,312
/ PRIOR FILING DATE: 2000-02-04
/ PRIOR APPLICATION NUMBER: US 60/207,456
/ PRIOR FILING DATE: 2000-05-26
/ PRIOR APPLICATION NUMBER: US 09/632,366
/ PRIOR FILING DATE: 2000-08-03
/ PRIOR APPLICATION NUMBER: GB 24263.6
/ PRIOR FILING DATE: 2000-10-04
/ PRIOR APPLICATION NUMBER: US 60/236,359
/ PRIOR FILING DATE: 2000-09-27
/ PRIOR APPLICATION NUMBER: PCT/US01/00666
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00667
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00664

```



```

; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.8
; OTHER INFORMATION: EST_HUMAN HIT: AW95346.1, EVALUE 0.00e+00
; OTHER INFORMATION: SWISSPROT HIT: P52739, EVALUE 0.00e+00
; OTHER INFORMATION: EST_HUMAN HIT: AW95346.1, EVALUE 0.00e+00
IS-09-864-761-37152

```

Query Match	5.6%	Score 297.5;	DB 10;	Length 515;
Best Local Similarity	21.0%;	Pred. NO. 1e-07;		
Matches 159; Conservative	73;	Mismatches 235;	Indels 291;	Gaps 29;

QY 226 GTGAASSTYPLPLEFSDIPKAQCGCKTAASSSSSSSSSSGAEPKQAFPHLYHPLGSOHPS 285
 Db 9 GEGGA-----MLF-----GRTEKKTLAGFAS-----RPPQR-----QPVSSNN--- 40
 QY 286 VGVGSGSHKPTPAPSPALPSPSTOOLIASPHLAFCTGTBOLLQAOCIGAANGLEMAASFGLL 345
 Db 41 -GLRVELEASPAOS-GNPEETDILKRIEVLGFQTVN--CGEC-----GLFSFKMTNLT 91
 QY 346 KPKN-GSGE---LGYEVYISLLEKPGGRHK-----CRCAKVPGSALQIHLR 390
 Db 92 SHORLHSGEKPYVCVGCEKGFSLKSLARHQKAHSGEKPYRCREGRGKRNKSTLIIHER 151
 QY 391 SHTGERPYKCNVCNCGNFFTRNGNLKVHFHNRHEKYRHWOMDHPVDEHLYITSSGLPYG 450
 Db 152 THSGEKPYCSCBGRGFSOKSNLIIHORTHSGEKP-----YCRECGKGPS 197
 QY 451 MSVPREKAEFEAGTPCGGYERKPLVASTALSALESILLTGTGTAVAGLPTFNKPYL 510
 Db 198 OKSANVRRHRT-----HLEKTIYVSDCLGFSRDSRLNIS----- 232
 QY 511 MKAVEPKSKADENTPPGSEGSALAGVADSGSATPMOLSKITVSLPSMALITNLKSTGSF 570
 Db 233 -----HORTHSGEKPYA---CKEGRCFRQRTT-----LVNH-QRTHSK 267
 QY 571 PFPLYLEPLGASRSEFSKIQOLVEKIDROGAVAVASTASCAPTTSAPAPSSASAGBNOCV 630
 Db 268 EKPYCGVCGSHFSFONSTL-----ISHRRITGT-----EKPYCG 302
 QY 631 ICLRLVSCPRALRLIHGOHGERPFKVCYGRAPSTGTGNLBAHFVGHKTSAPRAKONSP 690
 Db 303 VCGRGFSLSKSHLNRRQNHSHSGEKPYICKCKCGKGFSSQSNL-----IRHORTHSGEKPMVCG 358
 QY 691 ICQKFTNAVTLQOHVHMLGQIIP-----NGSALSDEGGAOENSSBOASTASGPFPO 746
 Db 359 ECGRGFSOKSNLVAHQRTH-SEGRPYVRCRGKGFSHQAG----- 397
 QY 747 PQSQPSPREEMSEDEEEDTDTDESLAGRGSESGEKALIVAGDSEEVGAEBE 806
 Db 398 -----LIRHKRKSRREKPYMCROGL-GFGNKS----- 424
 QY 807 VATSVAAPTVKEMDSNEKAPONTLPPPPRPDNLIDHPMEQSTDVSGAMEEBAKLEG 866
 Db 425 -----ALITHKRAHSEK----- 437
 QY 867 ISSPMALITOBEGESTPLVEELINLPEAMKKDRGESSGRACEVGGSFPTOTALEEHQ 926
 Db 438 -----PCV-----CRECGGFLQKSHLTLQHM 459
 QY 927 THPKDPLFTCVFCRGGFLDRATLKKN--MLLAHQVR 962
 Db 460 THTEKRP-YVCKTCGRGFSLSKSHLSNRKTTSYVHNRLP 496

RESULT 10
 US-09-764-864-883
 Sequence 883 Application US/09764864
 Patent No. US20020132753A1
 GENERAL INFORMATION:
 APPLICANT: Rosen et al.
 TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
 FILE REFERENCE: PT23
 CURRENT APPLICATION NUMBER: US/09/764.864
 CURRENT FILING DATE: 2001-01-17

```

; Prior application data removed - consult PALM or file wrapper
;
; NUMBER OF SEO ID NOS: 1792
; SOFTWARE: PatentIn Ver. 2.0
; SEO ID NO 883
; LENGTH: 577
;
; TYPE: prf
; ORGANISM: Homo sapiens
; OS-09-764-864-883

```

[illegible]


```
FILE REFERENCE: DEX-0291
CURRENT APPLICATION NUMBER: US/09/989,920
CURRENT FILING DATE: 2001-11-21
PRIOR APPLICATION NUMBER: 60/252,500
PRIOR FILING DATE: 2000-11-22
NUMBER OF SEQ ID NOS: 284
SOFTWARE: PatentIn version 3.1
SEQ ID NO 112
LENGTH: 610
TYPE: PRT
ORGANISM: Homo sapiens
US-09-989-920-212

Query Match
Best Local Similarity 24.1%; Pred. No. 4.8e-06;
Matches 90; Conservative 45; Mismatches 119; Indels 120; Gaps 10;

364 EKPCGRHKRCRCATVFGSDSALQIHLRSHTGERPKYCNVCNFTRGNLKVFHHRREK 423
132 ERP---YVCDVCGKTFERNNAGLKVHRLHTGERPKYCDVCGKAYISRSLKNHKGITLGE 188
424 YP-----HVQMNPHVPREHLDYVITSSGLPYGMSV---PPEK 457
189 KPYKSCYCEKSFNYSALAEQHKRIHTREKPFQDECGKAFRNNSGLKVHKRIHTGERPK 248
458 AEEBAGCPGGVERKPLVASTALNTESLTSLTSTGTAIVAGLPFNKPFVLMKAVEPK 517
249 CEE-----CGKAYISLSLINHKSVMHG-----EKP 274
518 SKADENPPGSEGAIVAGVADSGSATRMQLSLVTSPLPSWALLTNHLKSTGSPFPYVLE 577
275 FKCD-----CEKAI-----TYRITLNH-KKYHLGKPKCD 306
578 PLGASPSETSKLQQLVEKIDRGAAVAVASTAGAPTTAPAPSSASGPNOCVITCLRVLS 637
307 VCEKSFNTSLSLQHRVHTRE-----KPYECDRCCEKVER 341

638 CPRALRLHYGQHGRRPFCCKYCGRAFSTRGNLRAHFVGHKTSAPARAQNSCPICQKKET 697
342 NNSSLKVKHRIHTGERPKYCDVCGKAYISHSL---INHKSTHPKTPHTCDECGKAF 397

698 NAVTLOQHVRMLG 711
398 SSRLLSHKRVHLG 411

RESULT 12
US-09-764-864-976
Sequence 976, Application US/09764864
Patent No. US20020132753A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: PT223
CURRENT APPLICATION NUMBER: US/09/764,864
CURRENT FILING DATE: 2001-01-17
Prior application data removed - consult PAM or file wrapper
NUMBER OF SEQ ID NOS: 1792
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 976
LENGTH: 406
TYPE: PRT
ORGANISM: Homo sapiens
US-09-764-864-976

Query Match
Best Local Similarity 18.8%; Pred. No. 1.6e-05;
Matches 107; Conservative 52; Mismatches 150; Indels 259; Gaps 15;

364 EPPGGRHKRCRCATVFGSDSALQIHLRSHTGERPKYCNVCNFTRGNLKVFHHRREK 423
98 ERP---YECPECGKAFSEKSRLLKHORTHTGEEKPKYCDGCDKAFSAKSGRLHORTHTGE 154
```

```
424 YPHVQMNPHVPREHLDYVITSSGLPYGMSVPRPEKAEDEAGTPGGVERKPLVASTTALA 483
155 KP---FECHCECKSFNY----- 168
484 TESTLTLSTGSTAVAPAGLPFNKPFVLMKAVEPKSKADENTPPGSEGAIVAGVADSGSAT 543
169 -KSILIVHQRTHTG-----EKFECECKSFMSHG----- 199
544 RMQLSKLVTSLPSWALLTNHLKSTGSPFPYVLEPLGASPBETSKLQQLVEKIDRGAVA 603
200 -----LRNH-RRTHTGERPKCDECG---KAFKLSGLRKNHR----- 233
604 VASTAGAPTTAPAPSSASGPNOCVITCLRVLSCPRALRLHYGQHGRRPFCCKYCGRA 663
234 -----THTEKPKYCNVCNFTRGNLKVFHHRREKPFQDECGKAFRNNSGLKVHKRIHTGERPK 277
664 FSTRGNLRAHFVGHKTSAPARAQNSCPICQKKFTNAVTLQOHVHMLGQIIPNGSALSE 723
278 FSGKSLRVH---HRTHTGKPY-QCEECGKTFROKSNLRGHQRTHTG----- 321
724 GGAOENSSGOSTASGGSFPQPOSOQPPPEEMSEEEDEDEEDVTDEDLSLAGRS 783
322 -----EKPVECNCEGKAFSEKSVLKKHOR 345
784 ESGGEKALSVRGDSEEVYSGAEDEVATSVAPRTYKEMDSNEKAPQHTLPPRPDPNLDH 843
346 THTEK-----PYNCH- 356
844 PPEMGGTSDVSGAMEEAKLEGISSPMALLTOGEGTSTPLVELNLPEAMKKDPGESS 903
357 -----QCEGAFS---QKSNLRVHORTHTGKXP 380
904 GKACEYCGSFPTQALDEHOKTHPKD 931
381 YK---CDKGRFTFSOKSLREHOKAHPGD 406

RESULT 13
US-09-864-761-36336
Sequence 36336, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: Aeomica-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
```


PRIOR APPLICATION NUMBER: PCT/US01/00663
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00662
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00661
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00670
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: US 60/234,687
 PRIOR FILING DATE: 2000/09-21
 PRIOR APPLICATION NUMBER: US 09/608,408
 PRIOR FILING DATE: 2000-06-30
 PRIOR APPLICATION NUMBER: US 09/774,203
 PRIOR FILING DATE: 2001-01-29
 NUMBER OF SEQ ID NOS: 49117
 SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
 SEQ ID NO 36336
 LENGTH: 375
 TYPE: PRT
 ORGANISM: Homo sapiens
 FEATURE:
 OTHER INFORMATION: MAP TO AC011451.2
 OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.4
 OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.4
 OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.8
 OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.3
 OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.2
 OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.5
 OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.7
 OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.96
 OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.4
 OTHER INFORMATION: EST. HUMAN HIT: AF150229.1, EVALUATE 1.00e-102
 OTHER INFORMATION: SWISSPROT HIT: Q14585, EVALUATE 1.00e-124
 US-09-864-761-36336
 Query Match 4.6%; Score 244.5; DB 10; Length 375;
 Best Local Similarity 23.7%; Pred. No. 2.5e-05;
 Matches 89; Conservative 40; Mismatches 141; Indels 105; Gaps 11;
 QY 371 KCRFCAGVCGSDALQILHLSHNGEPRKCVNGRFFTRGNLKVHFRHREKYPHYQMN 430
 DB 39 ECHECGKAPVDHSSLKSHIRSHGSKPYOCECKCGKAFHFLACFFKKHMKTPTEE-----K 92
 QY 431 PPRVPEHLLVYITSSGLPYGMSVPEKAEKEAGTPGGGVGERKPIVASTALSTESTILL 490
 DB 93 PEECEECTAFSCSSSPFRAMKIKIGTNYCECKEKG-----FSCSSSLT-- 138
 QY 491 STGTSTAAVADGLTEFNKEVLMKAVEPKSKADENTPPGSEGSIAGYA-----DSGSATRM 545
 DB 139 -----EHKRIHSGDKPEYCKECKGKAFSCSSSLSKHKRIHSGDKP-Y 178
 QY 546 QLSKLVTSLSMALLNTHLK-STGSPFPFPPYLEPLGASPSFETSKLOQLYEKIDRGCAVAV 604
 DB 179 ECKECKGKAPSSSHLLIHNRIHNGE-KPEYCKECKGKAFSESSKL----- 221
 QY 605 ASTASGAPTTSPAPSSASGPNOQVYICLVSCPRALRIHNGHGG----- 651
 DB 222 --TVHGRHT-----GEKPYKCKCKGKAGYLPSSLSIHKRIHNGEPRYCECKGKAF 271
 QY 652 -----ERPRCKYKCGRAFPSTRGNLRAHFVGHKTSPPARAONSCPTQOKP 696
 DB 272 YLPTSLNTHVKKNOSREKPEYCKECKGKAFSCPSFRAHVDH-----TGKIYCKECKGKTF 327
 QY 697 TNAVTLOOHVRMLG 711
 DB 328 SRSSSLTEHLRTHSG 342
 RESULT 14
 US-09-764-864-1061
 Sequence 1061, Annomax Sequence Listing Engine vers. 1.1
 Patent No. US20020132753a1

```

: GENERAL INFORMATION:
: APPLICANT: Rosen et al.
: TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
: FILE REFERENCE: PT23
: CURRENT APPLICATION NUMBER: US/09/764,864
: CURRENT FILING DATE: 2001-01-17
: Prior application data removed - consult PALM or file wrapper
: NUMBER OF SEQ ID NOS: 1792
: SOFTWARE: Patentln Ver. 2.0
: SEQ ID NO 1061
:
: LENGTH: 513
:
: TYPE: PRT
:
: ORGANISM: Homo sapiens
:
: FEATURE:
:
: NAME/KEY: SITE
: LOCATION: (176)
: OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
:
: NAME/KEY: SITE
: LOCATION: (231)
: OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
:
: NAME/KEY: SITE
: LOCATION: (232)
: OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
:
: NAME/KEY: SITE
: LOCATION: (233)
: OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
:
: NAME/KEY: SITE
: LOCATION: (248)
: OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
:
: NAME/KEY: SITE
: LOCATION: (284)
: OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
:
: NAME/KEY: SITE
: LOCATION: (324)
: OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
:
: OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
:
: US-09-764-864-1061

```

Query Match	4.68;	Score 244.5;	DB 10;	Length 513;
Best Local Similarity	18.58;	Pred. No. 3.5e-05;		
Matches 131;	Conservative 77;	Mismatches 258;	Indels 241;	Gaps 21;
QY 282	HPFSVGVRGSRHKEPTTPAPSPALPGSTDDLIASP-ILAFPGTGLLAQCL--GAARGL--	336		
Db 12	HPYDMGGQSPRQHSVDLIGHEGVPTAESPLDCECKTFRGNRNDLIQRQIVHTGEASIMCD	71		
QY 337	---EAAASPELLPKPN-----GSGELGVEGVISLSEKPCGRHKKCRF	374		
Db 72	DCGKTSPQNSVNLKRNHSHSEKAYQSCGCGKAFRGHSDPFSRHQSHSSSRP---YMCNE	128		
QY 375	CAKVFSGSDSLQILIRSHTEGEPKYKCNVCGNFTFRGNLKVHNRHREKRPHYOMNPRV	434		
Db 129	CGKAFQNSNLKKNHOKSHMSEKPYECNCGCGAFRRSSNLIQRIHSHXEP-----	179		
QY 435	PEHLDVYITSSGRLPYGMSVPERKAEDEEAGTRPGGVERKRPVLVASTALSTATESITLLSTGT	494		
Db 180	---YVCECGCAFRRS---SNLIKHNRTHTG---EKRFEGEGCGKAFSSQSAHL-----	223		
QY 495	STAVAPGLPTFNKFLMKAVEPKSKADNPFGSGSALAGVADSGSATRMQLSKLVTSL	554		
Db 224	-----RKHQVHXXXXKRYECDCGKPRFSKX-----	249		
QY 555	PSWALLTNHLKSTGSPRPYVLLEPLGASPSFTSKLQ--LYEKLDRQGAVAVASTA--SGA	611		
Db 250	---NLIKHNRVHNGE--KPKYCSGCGKAFSSQSSSLIQRIHITGTEKPRHVCNCGKAFSFS	304		
QY 612	PTTSAAPAFSSASAGPMOCVCLRLVGLSPRALRLYHNGGGHEGEPFKCVKYGRAFFSTAGNLR	671		
Db 305	SVLRKHQIILHTGKPRKSCXCGKAFSSSALLIQHOGVITGDDPRYCHCEGCKFTGRSSNL-	363		
QY 672	AHEVGKITSAPAAQNSCPIQCKKFTNAVTLLQOHVMMHGGQIPNGSALSSECGAAQEN	731		
Db 364	---ILHQRVHTGKEPRCECTEGCTGSSQSSSTLIQRIHNG-----	400		


```
QY 732 SSEOSTAGSGSFPOQSOQSPREEMSEEEDEEDVTDSDSLAGSGSEGEKAI 791
DB 401 ----- 400
QY 792 SVRGDSEEVSAEEVATSVAAPTTVKEMDSNE--KAPQHTLPPRRPNLDHPQMEQ 849
DB 401 -----LKPHECNOCGKAFNRS-----SNLIHQKVT 427
QY 850 GTSVDGAMEREAKLEGISSPMALQIPEGEGSTPLVEELNLEPMKKDDGESSGRKACE 909
DB 428 GERYPT-----CVECGKGF-----QSHLIOHOIHTGERPK--CS 463
QY 910 VCGSGFPTOTALLEHQKTHPKDGPLFTCVCFRCROGFLDRATLKKHML 956
DB 464 ECGAFSGRSVLIQHRIHTGVKP-YDCAACGKAFSGRSKLINHQLI 509

RESULT 15
US-09-864-761-33476
; Sequence 33476, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263,6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 33476
; LENGTH: 439
; TYPE: PRT
; ORGANISM: Homo sapiens
```

```
; FEATURE:
; OTHER INFORMATION: MAP TO AF118808.1
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 7.1
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2.4
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.4
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2.4
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 6.7
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2.6
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 3.9
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.4
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 5.4
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 3
; OTHER INFORMATION: SWISSPROT HIT: P52742, EVALUE 0.00e+00
; OTHER INFORMATION: EST_HUMAN HIT: BE885790.1, EVALUE 1.00e-116
; OTHER INFORMATION: EST_HUMAN HIT: AF150239.1, EVALUE 1.00e-115
US-09-864-761-33476

Query Match 4.6%; Score 242.5; DB 10; Length 439;
Best Local Similarity 22.0%; Pred. No. 3,7e-05;
Matches 117; Conservative 42; Mismatches 201; Indels 173; Gaps 19;

QY 227 TGAASSTKPLPLPSPIKPKNOT---GKTTASSSSSSSGAERPKQAFNHLNPLGSDHP 283
DB 3 TKGESQNTDLSF--KPLISEQTVLGKTRPKRIDENNE---TKQSF-----CLSD 48
QY 284 FSYGVGVSRSHKPTPAPRPALPGSTQDLIASPHLAFP-GTTGLAQCILGAAGLEMAASP 342
DB 49 NSV-----DIREVYVLSQSMF-----LTPHQAVPSGERPYMCVGC-GKCFG-----88
QY 343 GLLKPKNGSELGYGEVYSSLEKPGGRHKRCFCAKVPFGSDALSQILNLSHTGERPKCV 402
DB 89 -----RSSHLIOHORINTGEKP---YVCSGCKAFSSSVLSKRRINTGEKPYECNE 138
QY 403 CGNRTTRGNLKVHFNHREKXYHQMNPHRYBELIDVYITSSGLPYGMSVPREKAEDEA 462
DB 139 CGKAFRVSSDLQOHKHTGKRPHECLCRAKFTQLSHLIOHORINTG-----186
QY 463 GTPGGVERKPLVASTTALSTESITLTS---TGTSTAVALPIETFNKFKVLMKAVERK 517
DB 187 -----ERYVCPRLCGKAKNHSITVLSHQVHVG-----EKP 217
QY 518 SKADENTPPSGESGALAGVADSGATRMQLSKVTSLSPW-----ALLTNIL 564
DB 218 HRCNE-----CGKTFESVKKRTLQHORINTGENBPTYCSECGKAFSDRSVLLIOHN 266
QY 565 KTCGSFPE-----PYVLEPLASPESTSKLDQLVKIDR 598
DB 267 VHTGEKPYECSECGKTFESHSTLMNHERINTGEKPYACYECGAFYQHSHL-----IQH 320
QY 599 QGAVALVASTAGAPTTAPAESSASGPNQCVICTLVLSCPRALRYLHGQNGGERPFCK 658
DB 321 QRV-----HTGEKPYVCGECGHAFFARSRLIGHRIHTGEKPRQST 361
QY 659 VCGAFSTRGCLRHFVGHKTSPPARAONSCPTCOCKFTFNATVTLQOHVNRHLG 711
DB 362 ECGKAFSLKATLIVLHRTN---TGEKPYECNCGKAFQSYVLIQHORINTG 410

RESULT 16
US-09-925-301-1357
; Sequence 1357, Application US/09925301
; Patent No. US20020052308A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA106
; CURRENT APPLICATION NUMBER: US/09/925,301
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05882
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1694
```


SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1357
LENGTH: 580
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: SITE
LOCATION: (3)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (526)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-301-1357

Query Match 4.6% Score 242; DB 10; Length 580;
Best Local Similarity 23.7%; Pred. No.5.2e-05;
Matches 105; Conservative 54; Mismatches 178; Indels 106; Gaps 16;

QY 322 TGLAAOCLGAARG-LEAASPGLKPKNGSG-----ELGYGEVSS-----LEK 365
DB 182 TGEKAPDCSDSGKSFIMSHLQGHILRTNGESLHEMKECGRGFIHSTDLAVRIOTHRSK 241
QY 366 PGRHRCRFCAKYPGSGSALOILHLSHTGERPKYKCNVCNRRFTTGNLKVHHRHREKYP 425
DB 242 P---YKCEKGKGRYSAYLINIHMGTHGTGDNPECKEKGKAFRSCOLQH-----RKT 292
QY 426 HYQMNHPVPEHLDYVITSSGLPYGMSVPRPEKAEAEAGTPGGVEKKPLVASTTALSATE 485
DB 293 HNGEKRYKCCDCGRAPFVSSCLSOHMKIHVGKPYECKECG-----IAFRSSOLTE 344
QY 486 STLLSTGTSTAVAPGLPTENKFLMKAVEPK--SKADENTPEGSGSAI-AGV-----A 537
DB 345 HLK-----FHTADP-----FECKICKGKSPFNSSCLSDHFRHTGHIKPKYKCK 386
QY 538 DGGSATRMQLSKVITSPSMALLTNHLKSTGSPFPYVLEPLGASGSEMSKL----- 589
DB 387 DCGKATTONSD-----LTKHAN-THSGERPVECKEKGKAFARSRLSEHRTHT 434
QY 590 -QOLVEKIDROGAVAVASTASG-----APTTSAPARSSASGP 626
DB 435 GKPFECVACGRKFAISSNLGHLRIHNGEKPEFLECGKATTHSSLNMMHTSHAKKP 494
QY 627 NCCVLCRLVSCPRLRLHYGQHGGERPKKVCGRAFSTRGNLRAHFVGHKTSPPARAQ 686
DB 495 FTCMEGKAFKFPCTCVNLHMRIHNGEKPKYKCKXCGKSFYSNSFOLHERTH-----TGEKP 550
QY 687 NSCPIQCKKFTNAVTLQOHVYMH 709
DB 551 YECKECGKAFSSSSSFRNHERRH 573

RESULT 17
US-09-735-367B-3
Sequence 3, Application US/09735367B
Patent No. US20020151477A1
GENERAL INFORMATION:
APPLICANT: Gustafsson, Jan-Ake
APPLICANT: Caixa, Francoise
APPLICANT: Antousson, Per
TITLE OF INVENTION: NUCLEAR RECEPTOR COACTIVATOR
FILE REFERENCE: 102093-100
CURRENT APPLICATION NUMBER: US/09/735,367B
CURRENT FILING DATE: 2000-12-12
PRIORITY APPLICATION NUMBER: US 60/174,544
PRIOR FILING DATE: 2000-01-05
NUMBER OF SEQ ID NOS: 18
SOFTWARE: PastSeq for Windows Version 4.0
SEQ ID NO 3
LENGTH: 2005
TYPE: PRT
ORGANISM: mammal
US-09-735-367B-3

Query Match 4.6% Score 242; DB 10; Length 2005;
Best Local Similarity 19.5%; Pred. No.0.00019;
Matches 225; Conservative 105; Mismatches 380; Indels 444; Gaps 46;

QY 19 AERGDASEEHHPQVCAKCAQFSDPTEFLAHQNSCTDP-----PYMVI-- 64
DB 777 AMGNASAGNHFSGHGMFNAFPGCAPN--GNQMSGQNGFVNNKDVTLSPLLVNLQ 834
QY 65 -----GGOENPSNSASAPRPEGHSRQYMDTEHSNPDGSSG-----PPDPTWG 111
DB 835 SDISAGHFGVNNKQNNANKPRKKRNSQODLTPTDRPAGLEBADQPLP--G 892
QY 112 PERGESSGQFLVAATGATAGGGGLIASPKIATPL--PPE--STRAPPPPP-- 163
DB 893 EOGTSLDNGCPKLEDFSRPPG-----YPSQVPEGRPLQGMPPQLMQHVAAPPQPPQQ 946
QY 164 -----PPPPVCGSHLNIPLILEELRLVQQRQIHQMOMTEQICRQVLLIGSGQT 214
DB 947 PDPQLPQQQQPPPP-----SOPQSQQQQQQQQQMMMLMQODPKS 987
QY 215 VGAPASPELPGTGAASATKPLPLFSPIKPAQTGKTAS--SSSSSSSGAEPKQAF-F 272
DB 988 VRLPVSONVHPPRG-----PLNPDQRMPMQOQSGSVPYMVSLQGPASVPPSPDKQMPM 1041
QY 273 HLXHPGSGQHPFSVGVGRSHKPTPAPAPALPGSTDQLASPHILAPGTGTLAAQCLGA 332
DB 1042 PVNTPUGS-----NSKKMYQESFQNPSS-----SP-----LEMASLPE 1076
QY 333 ARGLEAASBP-----LTKPKNGSGELGYGE-----VISLEKPGG 368
DB 1077 ASGSEAPSVPGGRNMPSHVVLPOQLMMWGPKPGRSPSLATAGATPOQRPVNSLSHG 1136
QY 369 RHKCRFCAKYPGSGSALOILHLSHTGERPKYKCNVCNRRFTTGNLKVHHRHREKYPHQ 428
DB 1137 HR---FPNVAAPQTQSRKTPKNRASPBY-----YEQTP 1167
QY 429 MNHPVPEHLDYVITSSGLPYGMSVPRPEKAEAEAGTPGGVEKKPLVASTTALSATE 488
DB 1168 NNRPPSTE-----PSEISLSPER-----LNASIAGLPPPOINT 1200
QY 489 LSTGTSTAVAPGLPTENKFLMKAVEPKSKADENTPGESEGSALAGVADSGSATRMQLS 548
DB 1201 PL-----PPRPNLRGFDQGLNPPT----- 1221
QY 549 KLVTLSPSMALLTNHLKSTGSPFPYVLEPLGASGSEMSKLOQVEKIDROGAVAVASTA 608
DB 1222 -----LKAIGQAPSNLTMPNSFATPQTHKLDSVVVNSGKQ-----SN 1259
QY 609 SGAPTSAPAPSSSASGPNOCVLCRLVLSCPRALRLHYGQHGGERPKKVCGRAFSTRG 668
DB 1260 SGATKRASPS-NSRRSSP-----GSSR----- 1280
QY 669 NLRAHFVGHKTSPPAARQNS-CP-ICOKKFTNAVTLQOHVYH-----HLGG 712
DB 1281 -----KTPSPGRQNSKAPKLTLASQTNAAAL-ONVELPRNVLVSPPTPLANPPVG 1330
QY 713 QIPNGS-----ALSEGGAGQENSSEOSTASGSPSPF---QSQQPSSEEMSEE 761
DB 1331 SFPNNSGLNPQNSTVSAVAVAGVEDNKESLNV-----PDSQOQNSQKREQVNIET 1383
QY 762 EDEDEEEDVTDSDSLAGNGSESGEKAISVRGSEEVSGAEFEVATVVAAPTYKEMD 821
DB 1384 KAVPAQEVKVVVEEDSKKQGPSPDKPLPVEENKKNLVSPAMRE-----APTSLSQL 1437
QY 822 SNEKAPQHTLPPP-----PP-----PPDN----- 840
DB 1438 DMSGAPNVTIKPPGLDLEVTYPVAVSGEDLKASVITPLQDLSSEKPESNSINTLPHSNEL 1497
QY 841 ---LDHPQMEGTS---DYSGAMEEBAKLEGISSPMALQOEGGCTSPVLYEELNLPRA 894
DB 1498 CSSLVHPELSEVSSNAVPSLPPVMSRPVSSSISTPLPPOIVTFVTSNPITTSANTSA 1557
QY 895 MKKDPGESSGRKACEVCGOSFPT--OTALBEHQKTHPKDGLFTGCVFPROG----- 943


```
Db 1558 L-----PTHLOSALMSTVVTMPNAG---SKYVNSGQSAQSNAR 1554
Qy 944 -----FLDRATLKHHMLAHQVPPFAP-----HGPONI--ATLS 976
Db 1595 PQFTTFPFINSSIIQVKKSGQSPRTIPAAPLTNTSGLMPPSVAVGRLHPQNIKFFSSAP 1654
Qy 977 LVPGCSSSIPSPGL 990
Db 1655 VPPNALSSSPAPNI 1668

RESULT 18
US-09-735-367B-2
; Sequence 2, Application US/09735367B
; Patent No. US2002015147A1
; GENERAL INFORMATION:
; APPLICANT: Gustafsson, Jan-Ake
; APPLICANT: Caltra, Françoise
; APPLICANT: Antonsson, Per
; TITLE OF INVENTION: NUCLEAR RECEPTOR COACTIVATOR
; FILE REFERENCE: 102093-100
; CURRENT APPLICATION NUMBER: US/09/735,367B
; CURRENT FILING DATE: 2000-12-12
; PRIOR APPLICATION NUMBER: US 60/174,544
; PRIOR FILING DATE: 2000-01-05
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 2063
; TYPE: PRT
; ORGANISM: Human
US-09-735-367B-2

Query Match 4.6%; Score 242; DB 10; Length 2063;
Best Local Similarity 19.5%; Pred. No. 0.00019;
Matches 225; Conservative 105; Mismatches 380; Indels 444; Gaps 46;

Qy 19 AERGDASEHHNPQVCAKCAQPSDPTFLAHQNSCCTDP-----PVMVRI-- 64
Db 835 AMQNSASQNHFSGCHMSFNAFPGSAPN--GNQNSGQNPGEFVYKNDVTLTSPLLVNLIQ 892
Qy 65 -----GQENPNSNSASSAPRECHSRQVMDTEHSNPDGSSG-----PPDPTWG 111
Db 893 SPISAGHFVGNKKNNTNANKPRKKKKKNSQDLNTPDTTRPAGLEADQRPPLP--G 950
Qy 112 PERRGESSGQFLVAATGTAAGGGGLILASPKIGATPL--PPE--STRAPPPPP-- 163
Db 951 EGGISLIDNSGPKLPFEFSNRRPG-----YPSQRYEORPLQMPRLQMHVAAPRPQPOQ 1004
Qy 164 -----PPPPRGVSGHLNPLTLEELRVLDQROQHOMTEQIGROVILLGSLQGT 214
Db 1005 POPQLPQQQPPPP-----SOPQSQQQQQQQQMMMLMMQODPFS 1045
Qy 215 VGAPASPELPGGAASSTKPLPLFSPIKFAQTKGTAS--SSSSSSSSGAPPRQAF-F 272
Db 1046 VALLPVSONNHRPG-----PLNPDQRMQRQSGSVVVMVSLQRPASVPRPDKQRMNM 1099
Qy 273 HLVHPLGSHRFSVGVGRSHKPTPADSPALPGSTDOIILASPHLAFPGTTGILLAAQCLGA 332
Db 1100 PVMTPPLGS-----NSRKMYQGESPPQSS--SP-----LAEMASLPE 1134
Qy 333 ANGLEAAASRG-----LLKPKNGSGELGGE-----VSSLEKPPG 368
Db 1135 ASGSEAPSVYGGGNNMPSHVLLPONQLMTTPKPGSPPLSATQGTPOQRPVNSLPSSSHG 1194
Qy 369 RHKCRCAKFFGSDSLQILHLSHTGERPYKCNVCGNRFETTGNLKVHRRHREKYPHYQ 428
Db 1195 HH-----FPNVAAPRTQTRPKTPRNASPRPY-----YRQTP 1225
Qy 429 MNDPVPENHLDYVTSSGLPYGMSVPRPKAEBAEACTPGGCGVERKKPLVASI;ALSATESLT 488
Db 1226 NNRPSPTE-----PSEISLSPER-----LMSIAGLFPPOINI 1258
```

```
Qy 489 ILSTGSTAVAPGLPTEFNKRVLMKAVEPKSKADENTPPGSEGSALAGVADSGSATRMOLS 548
Db 1259 PL-----PPRNLNKRFFDQGLMPTT----- 1279
Qy 549 KLVTLSPWALLTNHUKSTGSPFPFVYLEPLGASPSSETSKLQQLVEKIDRGAVAVASTA 608
Db 1280 -----LKAIGQAPSNLTMTNPSNFATPQTHKLDYSVVVNSGKQ-----SN 1317
Qy 609 SGAPPTTSAPAPSSASGPNOCVICRLVSLCPRALRLHNGGCGEFPFKCYGGRAPSTRG 668
Db 1318 SGATKRAAPS-NSRRSSP-----GSSR----- 1338
Qy 669 NLRAHFVGHKTSAPAAKQNS-CP-ICQKKTNAVTLQOHVR-----HLGG 712
Db 1339 -----KTPSPGQNSKAPKLTILASQTMALL-QNVELPRVILSPTPLANPPVG 1388
Qy 713 QLPNGS-----ALSEGGAQENSSEOSTASGPGSFQPP---OSQQPSPEEMSEE 761
Db 1389 SFPNNSGLNPNQNSTVSVAAGVVEDNKESLNV-----FQSDSCQNSQKREQVNIEL 1441
Qy 762 FEEDDEEEEDVTDSDSLAGRSSESGEKAISVRGDSSEEVSGAEEEVATSVAPPTVKEMD 821
Db 1442 KAVPAQEVKNAVPEQSKDQGPSPDPNKLPSVEENKNLVSPAMRE-----APTSLSQL 1495
Qy 822 SNEKAPQHTLPP-----PP-----PPDN----- 840
Db 1496 DNGAPRVNTIKPGLTLDLEVTPPVYSGEDLKKASYIPTLQDLSSEKPEPNSLNLPHSNEL 1555
Qy 841 --LDHPQMEQGS--DVSAMEEFAKLEGISPPMAALTOEBGSTSPPLYBELWLEPA 894
Db 1556 CSSLVHPBELSEVSNVAFPIPVMSRPVSSSISTPLPNOITVFVTSNPIITSAVNTAA 1615
Qy 895 MKKDPGSSGKKAQEVGQSPPT--QTALEHQKTHPKDGPLTCVCFQCGO----- 943
Db 1616 L-----PTHLOSALMSTVVTMPNAG---SKYVNSGQSAQSNAR 1652
Qy 944 -----FLDRATLKHHMLAHQVPPFAP-----HGPONI--ATLS 976
Db 1653 PQFTTFPFINSSIIQVKKSGQSPRTIPAAPLTNTSGLMPPSVAVGRLHPQNIKFFSSAP 1712
Qy 977 LVPGCSSSIPSPGL 990
Db 1713 VPPNALSSSPAPNI 1726
```

```
RESULT 19
US-09-764-864-948
; Sequence 948, Application US/09764864
; Patent No. US20020132753A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT223
; CURRENT APPLICATION NUMBER: US/09/764,864
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1792
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 948
; LENGTH: 338
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (335)
; OTHER INFORMATION: xaa equals any of the naturally occurring L-amino acids
US-09-764-864-948
```

```
Query Match 4.6%; Score 241.5; DB 10; Length 338;
Best Local Similarity 24.2%; Pred. No. 3.1e-05;
Matches 88; Conservative 38; Mismatches 154; Indels 83; Gaps 11;
```


OY 545 MOSKIVTSLPSMALLTHMLK-STGSEFPYVLEPLASPSETSL--QQLVEKIDROGA 601
 Db 336 CNI-----YLHQRHTGE--RPVYCECGKSPCKSVLIDHQRVHTTGERPYE 382
 OY 602 VA---VASTASCAPPTSAAPSSASGPNQVCYICLVLSCPRAVLRLHGGHGERPFKC 657
 Db 383 CSBCKGVFSSSG---LFRHRAHRTKTKPECESECEKSFSCKTDLIRQTVHTGERPYEC 439
 OY 658 KVCGRAFSTGNLNAHFVGHKTSPARAONSCPICQKKFTNAVTLQOHVRLHG 711
 Db 440 SVCGKSPFR--KTHLIRHQTVHTNERPEYCECGKSVSSSALLHQRVHTG 489
 RESULT 25
 US-09-864-761-34333
 Sequence 34333, Application US/09864761
 Patent No. US20020048763A1
 GENERAL INFORMATION:
 APPLICANT: Penn, Sharon G.
 APPLICANT: Rank, David R.
 APPLICANT: Hanzel, David K.
 APPLICANT: Chen, Wensheng
 TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
 FILE REFERENCE: Aecomica-X-1
 CURRENT APPLICATION NUMBER: US/09/864,761
 CURRENT FILING DATE: 2001-05-23
 PRIOR APPLICATION NUMBER: US 60/180,312
 PRIOR FILING DATE: 2000-02-04
 PRIOR APPLICATION NUMBER: US 60/207,456
 PRIOR FILING DATE: 2000-05-26
 PRIOR APPLICATION NUMBER: US 09/632,366
 PRIOR FILING DATE: 2000-08-03
 PRIOR APPLICATION NUMBER: GB 24263.6
 PRIOR FILING DATE: 2000-10-04
 PRIOR APPLICATION NUMBER: US 60/236,359
 PRIOR FILING DATE: 2000-09-27
 PRIOR APPLICATION NUMBER: PCT/US01/00666
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00667
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00664
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00669
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00665
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00668
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00663
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00662
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00661
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00670
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: US 60/234,687
 PRIOR FILING DATE: 2000-09-21
 PRIOR APPLICATION NUMBER: US 09/608,408
 PRIOR FILING DATE: 2000-06-30
 PRIOR APPLICATION NUMBER: US 09/774,203
 PRIOR FILING DATE: 2001-01-29
 NUMBER OF SEQ ID NOS: 49117
 SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
 SEQ ID NO 34333
 LENGTH: 311
 TYPE: PRT
 ORGANISM: Homo sapiens
 FEATURE:
 OTHER INFORMATION: MAP TO AC008165.2
 OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.5
 OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.9

```

; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.97
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.2
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.7
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 2.3
; OTHER INFORMATION: EXPRESSED IN HEL100, SIGNAL = 2.5
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.7
; OTHER INFORMATION: SWISSPROT HIT: P52742, EVALUATE 1.00e-115
; OTHER INFORMATION: EST_HUMAN HIT: AUI33067.1, EVALUATE 3.00e-85
US-09-864-761-34333

Query Match      4.48; Score 234; DB 10; Length 311;
Best Local Similarity 23.18; Pred. No. 6.6e-05;
Matches 86; Conservative 44; Mismatches 133; Indels 110; Gaps 12;

358 EVISSLSEKPGGRHRCRCAKYFGSDSALOHLRSHTEGGERPYGCMGCGNRTTSGNLKVHF 417
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 21 EMHTHTGKPP--HCHCLCGKAFTHCSDLRKHERHTHGEKPYGCHLCKGAFSKSSNLRHE 77
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
418 HRR-REK-----YPHYQMPHPVPEHLIDYVITSSGLPYGMSVPEK-----AEE 460
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 78 MHTREKAQOICHLCKGAFTHCSDLRKHERTHL-----GDKRYGCLLCKGAFSKCYLRQ 131
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
461 EAQTPGGVGRKRPVASTLTALSTESLTLLTGTSTAVAPGLPTFNKPYLMKRVKRS-- 518
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 132 HERTHNG--EKP-----YCHCLCKGAFSHCSHL 157
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
519 KADENTPGSE--GSAIAGVADSGATRRMOLSKLVTLSLPSMALTLNHLKSTGFPFPYVL 576
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 158 RQHERSHNGKRPCHCLCKGAFTHCSDLRKHERHTG-----EKPYEC 200
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
577 EPLGASPESTSKLOOLVEKIDROGAVAVASTAGAPTTAPAPSSASGPNOCVILRYL 636
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 201 HVGKGAFTSSDLR-----HERHTHGEKPYEHLCKGAF 235
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
637 SCPRALRLHYGCGGEPPEKCKVCGRAFTSTGNLRAHFVGHKTSPPARAONSCPTQCKP 696
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 236 NHSSVLRHRRHTHTGGERPYECNICGKAFFNRSTYFRLHRRVH----TGKRPYVCPLCKGAF 291
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
697 TNNAVTLQOAHVRFH 709
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 292 SKFFNLROHERTH 304
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

RESULT 26
US-09-864-761-35894
; Sequence 35894, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aemica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30

```



```

    PRIOR APPLICATION NUMBER: PCT/US01/00664
    PRIOR FILING DATE: 2001-01-30
    PRIOR APPLICATION NUMBER: PCT/US01/00669
    PRIOR FILING DATE: 2001-01-30
    PRIOR APPLICATION NUMBER: PCT/US01/00665
    PRIOR FILING DATE: 2001-01-30
    PRIOR APPLICATION NUMBER: PCT/US01/00668
    PRIOR FILING DATE: 2001-01-30
    PRIOR APPLICATION NUMBER: PCT/US01/00663
    PRIOR FILING DATE: 2001-01-30
    PRIOR APPLICATION NUMBER: PCT/US01/00662
    PRIOR FILING DATE: 2001-01-30
    PRIOR APPLICATION NUMBER: PCT/US01/00661
    PRIOR FILING DATE: 2001-01-30
    PRIOR APPLICATION NUMBER: PCT/US01/00670
    PRIOR FILING DATE: 2001-01-30
    PRIOR APPLICATION NUMBER: US 60/234,687
    PRIOR FILING DATE: 2000-09-21
    PRIOR APPLICATION NUMBER: US 09/608,408
    PRIOR FILING DATE: 2000-06-30
    PRIOR APPLICATION NUMBER: US 09/774,203
    PRIOR FILING DATE: 2001-01-29
    NUMBER OF SEQ ID NOS: 49117
    SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
    SEQ ID NO: 35894
    LENGTH: 340
    TYPE: PRT
    ORGANISM: Homo sapiens
    FEATURE:
    OTHER INFORMATION: MAP TO: AF166490.1
    OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 5.5
    OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 4.5
    OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 3.8
    OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 3.6
    OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 9.4
    OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2.3
    OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 3.6
    OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 6.1
    OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 4.6
    OTHER INFORMATION: ESTIMATED IN HEART, SIGNAL = 3
    OTHER INFORMATION: EST_HUMAN HIT: AF150239.1, EVALUATE 1.00e-91
    OTHER INFORMATION: SWISSPROT HIT: P52742, EVALUATE 1.00e-118
    US-09-864-761-35894

Query Match 4.48; Score 234; DB 10; Length 340;
Best Local Similarity 23.18; Pred. No. 7.2e-05;
Matches 86; Conservative 44; Mismatches 133; Indels 110; Gaps 12.

QY 358 EYVSSLEKGGGRKRCRCACVAFGSDSALQIHRSHTGEPYKCVGCGNRPTTGNLKVHF 417
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 50 EIMIHGKEP--HGCHLCGKRAFTHCSDLRKHERHTGEPYCGCHLCKAFKSSNLRHE 106
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 418 HHH-REK-----YPHQMNPHVPEHNDYVITSSGLPYGMSVPERK-----AAE 460
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 107 MHTTRKKAQIHLCKGKRAFTHCSDLRKHERHTL-----GDKPYGCLLGKAFKCSYLRL 160
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 461 EAGTPGGVEKRPDLVASTALSTATESITLLSTGTSTAVAPGLTFENKFLMKAVEPKS-- 518
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 161 HERTHNG---EKP-----YECHLCKGKAFSHCSHL 186
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 519 KADENTPPGSE--GSAIAGVADSSATPMQSLKLYTSLPSMALLTNHLKSTGSEPPFPYVL 576
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 187 ROHERSHNEKRPCHLCGKRAFTESSYLKRHERHTG-----EKEPYEC 229
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 577 EPLGASPESETSKLOQLVEKIDRGAVAVASTASGAVTSPAPAPSSASGPNOCYICLRVL 636
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 230 HVCCKAFTESSDLR-----HERTHNGEKPYECHLCKGKAF 264
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 637 SCPRLALRYHGSGGERPFKCKVCGRAFTSGNLRHAFVGHKTSPPARAONSCPIQCKKF 696
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 265 NHSSVLRHRERTHTGKRPYECNICGKAPNRSVNFRLRRHVH-----TGEKPYVCCILGKAF 320
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 697 TNAVTLQGNVRHM 709

```

```

Db          321 SKFNLRHERTH 333
:          :||| | |
RESULT 27
US-09-864-761-38000
: Sequence 38000, Application US/09864761
: Patent No. US20020048763A1
: GENERAL INFORMATION:
:   APPLICANT: Penn, Sharon G.
:   APPLICANT: Rank, David R.
:   APPLICANT: Hanzel, David K.
:   APPLICANT: Chen, Wensheng
: TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
: FILE REFERENCE: Aecmics-X-1
: CURRENT APPLICATION NUMBER: US/09/864,761
: CURRENT FILING DATE: 2001-05-23
: PRIOR APPLICATION NUMBER: US 60/180,312
: PRIOR FILING DATE: 2000-02-04
: PRIOR APPLICATION NUMBER: US 60/207,456
: PRIOR FILING DATE: 2000-05-26
: PRIOR APPLICATION NUMBER: US 09/632,366
: PRIOR FILING DATE: 2000-08-03
: PRIOR APPLICATION NUMBER: GB 24263,6
: PRIOR FILING DATE: 2000-10-04
: PRIOR APPLICATION NUMBER: US 60/236,359
: PRIOR FILING DATE: 2000-09-27
: PRIOR APPLICATION NUMBER: PCT/US01/00666
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00667
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00664
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00669
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00665
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00668
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00663
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00662
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00661
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00670
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: US 60/234,687
: PRIOR FILING DATE: 2000-09-21
: PRIOR APPLICATION NUMBER: US 09/608,408
: PRIOR FILING DATE: 2000-06-30
: PRIOR APPLICATION NUMBER: US 09/774,203
: PRIOR FILING DATE: 2001-01-29
: NUMBER OF SEQ ID NOS: 4917
: SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
: SEQ ID NO 38000
: LENGTH: 983
: TYPE: PRT
: ORGANISM: Homo sapiens
: FEATURE:
: OTHER INFORMATION: MAP TO AL031277.1
: OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.6
: OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.4
: OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.1
: OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2.6
: OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.8
: OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.98
: OTHER INFORMATION: EST_HUMAN HIT: AJ0140221.1, EVALUATE 1.00e-111
: OTHER INFORMATION: SWISSPROT HIT: Q61116, EVALUATE 1.00e-09
US-09-864-761-38000

```


Best Local Similarity 19.1%; Pred. No. 0.00023;
Matches 187; Conservative 100; Mismatches 383; Indels 307; Gaps 35;

OY 1 MAQETSSSLGFCG-EPARGGDASEHHPOYAKCAQGFDPHEFLAHQNCCTDPP 59
Db 109 MSKEMTASSAFSSVYCNODPLDSSGVKOK--AEGTKPTVQKSVLIDSVYKHKCQDSE- 165
OY 60 VMVILIGQENPNSN-----SASAPRPPGHSRSQVMDTEHS-----NPPDSSGSGP-- 105
Db 166 -----GKEFESHSHVQPTCSAVKKRKPPTCMLOKYLINNEYNGIDLVEKNAD-CTRSPSP 219
OY 106 -----PDPPTWPERKGEESGQFLVAATGTAAGGGGILLASPKIGATPLPP--EST 155
Db 220 CKSLEAOPDPLDGP-----GSGFPAPVTESTPDVCPSSPALQTPSLSSQGLPPLIPTD 273
OY 156 PAPPRRPP-----PPRPGVSGSHLNTPLILEELRVLDQGRQIHQNMOTDQICQVILLG 209
Db 274 PSSPPPCPPVLTVAATPPPLPLTVPLPAP----- 302
OY 210 SLGQTVGAPASPELPP--GTGAASSTRPL-----PLFSPIKP-----AQTGKTAS 254
Db 303 -----SSASHPHCSPPLSNATQSPRLPLISPIVSPSPSPPIRPERLMSAASGPPTLS 356
OY 255 SSSSSSSSGAEPKQAFPHLYHPLGSHQHPFSYGVGGRSHKPTTPAPSPALD--GSTDQLI 311
Db 357 SSSSSSSSSS-----SFSSSSSSSSPSPPLSAISSVSSGDNLE 396
OY 312 AS-PHILAPFTTGLLAQCLGARGLEAAASPGILKPKNGSGELGGEVYSSLEKPPGRH 370
Db 397 ASILPMISFK-----QELENEG-LKPREPOSAADQDVVQ--ETFNKNF 438
OY 371 KCRFCACVFEGSDSALOIHLSRHTGERPYKCNVCGNRFTRGNLKVH----- 416
Db 439 VCVNCSPLFSLINDLTHLSHAEMEPKCEFCVOLFKKTDLSHRFLHGVGNITVCS 498
OY 417 -----FHRH-PEKPHVQNMNH-----PVREHLDYV----- 441
Db 499 VCKKEAFILCNLOHODLHPDKVCTHHEFESGTLRQNFPTDPSKAHVEHQSLPEDPLE 558
OY 442 -----TSSGLP-----YGSVPERKEEKEGTTGGG 468
Db 559 TSKEEBELNDSSELYTTIKTMASGIKTKDPDVRILGNQHYBSEKPPFOYHNHRPMGIG 618
OY 469 VERKPLVASTTALATESLTLSTGTSTAVAPGL-----PTFNKFLVM 511
Db 619 VYATNFTTHNIPOFTTAIRCTKCGKGVDMPELKHILLACASASDKKRTPKKNPPLK 678
OY 512 KAVEPKSKADENTPPGSEBSAIAVADSGSATRMQLSKLVTSILPSWALLTNHLKSTGSPF 571
Db 679 QTVQPK-----NGVVVDNDSGNKAFRRMGOPKRLNFSVELSKMSNKKLIKLNALKKKQOLV 733
OY 572 FRYVLEPLGASPFETSKIQOLVE-----KIDRQAVANASTASGAPTT 614
Db 734 QKAILQK-NKSAKOKAKADKNACESSSHICPCNKEFTYIGSLKHAAFSCPKKPLSPKK 792
OY 615 -----SAPAPSSASGPN-----OCVICLRVLSCPRALRILHYGOHGERPPKC 657
Db 793 KYSHSKKGGHSSPASSDKNSNHRRTADAEIKMQSMQTP---LGKTRASSGPTQV 848
OY 658 KYCGRAPSTRGNLR-AHFVGHKTSPPARAQNSCPICQKKFTNA----- 699
Db 849 PLPSSSFRSKQNVKFAFASVSKKPKSSSSILRNSSPIRMAKITHVEGKPKAVAKNHSQOLS 908
OY 700 --VTLOOHVMMHIGQOIPNGSALSSEGGGAQENSEQSTAGSPGSPFOQSOQSPSEEE 757
Db 909 SKTSRSLHVRVQSKAVALQSKSTLASKKRTDRFNKTSRERSGGPVY---RSLQLAAAD 964
OY 758 MSEEDEEEDDEEDVTD 774
Db 965 LSENKREDSAKQELKD 981
RESULT 28

US-09-764-864-893
; Sequence 893, Application US/09764864
; Patent No. US20020132753A1

; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT23
; CURRENT APPLICATION NUMBER: US/09/764,864
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1792
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 893
; LENGTH: 361
; TYPE: PRP
; ORGANISM: Homo sapiens
US-09-764-864-893

Query Match 4.4%; Score 232; DB 10; Length 361;
Best Local Similarity 20.8%; Pred. No. 9.6e-05;
Matches 73; Conservative 33; Mismatches 115; Indels 130; Gaps 8;

OY 364 EKPGRHKCRFCANVFEGSDSALOIHLSRHTGERPYKCNVCGNRFTRGNLKVFHHRREK 423
Db 121 EKP---HKTECGKSFNEKSTLIVHQRTHTGKPYECDVCGKTFQKSNLGVHQRTHSGE 177
OY 424 YPHYQNMHPVREHLDYITSSGLPYGMSVPEKAEAGTPGGGVKRPVASTLSA 483
Db 178 KPEECNECEKAFSOKSYTML-----HQKHNTG---EKPEECNECEKAF 217
OY 484 TESTLTLSTGTSTAVAPGLPTFNKFLVMKAVEPKSKADENTPPGSEGSATAGVADSGSAT 543
Db 218 SQKSYLL-----IQHRTTEK----- 234
OY 544 RMQLSKLVTSILPSWALLTNHLKSTGSPFPPIVLEPLGASPFETSKL---QOLYEKIDRQ 600
Db 235 -----PYKCNCEKAFREKSKLIIHQRI----- 257
OY 601 AVAVASTASGAPTTAPAPSSASGPNOCVICLRVLSCPRALRILHYGOHGERPPKCYC 660
Db 258 -----HGEKPYECPVCKWAFSQKSLIIRHQRTHGKRYACTEC 297
OY 661 GRAFSTRGNLRAHFVGHKTSPPARAQNSCPICQKKFTNAVTTLOOHVMMHIG 711
Db 298 GKAFREKST---FTVHQRTHTGKPYKCTECGKAFQKTSNLIVHQRTHAG 344

RESULT 29
US-09-764-864-954
; Sequence 954, Application US/09764864
; Patent No. US20020132753A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT23
; CURRENT APPLICATION NUMBER: US/09/764,864
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1792
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 954
; LENGTH: 708
; TYPE: PRP
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (328)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-864-954

Query Match 4.4%; Score 231; DB 10; Length 708;
Best Local Similarity 19.2%; Pred. No. 0.00021;
Matches 181; Conservative 88; Mismatches 328; Indels 348; Gaps 39;


```
OY 108 PTWGBRGESESSGQFLVATGTAAGGGGLILASPKLGATPLPEESTPAPPPPPPP 167
| | | | | : | : | : | | | | |
Db 13 PLMHGRCGTSSAAMNASEGESFAGS---VOIP--GGTIVVELTP----- 55
OY 168 PGVSGHLNPLILEELRVLQQRQIHOMQMTQEQICRQVLLLSIGQTVGAPASSELPGT 227
| | | | | : | | | | |
Db 56 -----DIIHC---GICKQ--QFNLDARVANHKGCGQLTGT 86
OY 228 GAASSTKPLPLFSPKPAQOTGKTASSSSSSSSSGAEPKQAFH---LYHPLGSH-- 282
| | | | | : | : | : | : | : | : |
Db 87 SAAASTVOEVSSEETPAVOTQTTTITSETQITIVSABEFVEHGYQTLPTESNENO 146
OY 283 ---PVSQVGVSRSHKPTPAVSPALPGSTDLIASPHLAPPGTGLLAAQCLGAARGLEAA 339
| | | | | : | | | | |
Db 147 TATVLSLAKSTKKTTPPAQ-----KRLNCCYPCGQ----- 179
OY 340 ASPLGLKPRNGSELGCGYFVLSLEKPGGRHNCRCFAKVFSGDSALQIHLRSHTERPYK 399
| | | | | : | | | | |
Db 180 ----FKTAYGKMDMRHLKIHGDKP--HKCEVCGKCFSRDKLKTMHRCCHTGVRKYK 231
OY 400 CNVGNRTTRGNLKVHPIHREKRP--HYQMANHPVPEHLDYVITSSGLPYGMSVPEKA 458
| | | | | : | | | | |
Db 232 CKTCYAAADSSSLKHLRIHSDERPFKCQICPY----- 265
OY 459 EEEAGTPGGGVERRKPLVASTTALSTESLTLSGTSTAVAPGLPTFN-----KFLMK 512
| | | | | : | | | | |
Db 266 -----AARNSQLTVHLR-----SHTGAP-----FOCWLSAKFKLSS 299
OY 513 AVEPKSKADENTPPGSEGSALAGVADSGSATRMQLSKLVSLPFWALLTNHLK---STGS 569
| | | | | : | | | | |
Db 300 DLKRMHVRHSGEKP-----FKCEFCNVRCMK--GXLSHLRIKHSGNN 341
OY 570 FPRPVLELGAISPSTSKLQOLVEKIDROGAVALAVASTAGAPTTSAAPSSSASGPNOC 629
| | | | | : | | | | |
Db 342 FKCPH-CDFLGGSKALRLRHRSRV-----HOSEHPEKC 372
OY 630 VICTLVLSCPRLLRLHYGONGGERPFCKVCGRAFSTR--GNLRANF----- 674
| | | | | : | | | | |
Db 373 SFCSSVSCSSKALRIHERIHCTRPFCNYC--SFDTKQPSNLKHKMKFKHDMYKTEL 430
OY 675 ---VGHKTSPLA---ARAQNSCPICQCKFTNAVTLQOHVHMLGQIPNGSALSSEG 725
| | | | | : | | | | |
Db 431 ERKDIQGRQSRQVAKLDKKSFFHCIDICDASFMRDSDLRSHKRON-----SEYSEK 481
OY 726 GA-----AQENSSQBSTASGSPFQPOSQOPSPREEEMSEEEDEEDVDEBGL 778
| | | | | : | | | | |
Db 482 NSDVTYLOFQIDPSKQAPATPLTVGHLOVP--LOPSOVPOFSE----- 521
OY 779 AGRGSESGEKAISVGDSEVSGAEDEVATSVAPTTV-----KEMDSNEKAP-- 827
| | | | | : | | | | |
Db 522 -----GRVITIV---GHQVQPAQNTIVQAAAAVNIIVPALVAQNPBELPGNSRLQIL 570
OY 828 -QHTLPRPPRPDNLDPORMEQGTSQVSGAMEBEAKLEGISSPMALTOBGEESTPLV 886
| | | | | : | | | | |
Db 571 RQVSLIAPROP-----SRCPSEAG-----ANTQPAVLITTH-EQDGAATL 609
OY 887 BELNLPEAMKKDPGESSGRKACEVCGOSFPTQTALEHQKTHPKDGLFTC-----VF 939
| | | | | : | | | | |
Db 610 HQTLLPETA-SGGQEGSG-----NOTFTTSSGI-----TCDFEGLNAL 647
OY 940 CRQGFIDRATLKHMMLAHQHPVPFAPHGPNILITLSLPGCCSSS 984
| | | | | : | | | | |
Db 648 IOEGTAETVVS-----DGGQNTAVATTAPVFFSS 677
```

```
APPLICANT: ZHOU, XIN
TITLE OF INVENTION: MASTER BONE FORMATION TRANSCRIPTION FACTOR:
FILE REFERENCE: UTXC:666
CURRENT APPLICATION NUMBER: US/09/734,329
CURRENT FILING DATE: 2000-11-30
NUMBER OF SEQ ID NOS: 6
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 2
LENGTH: 428
TYPE: PRT
ORGANISM: Mus musculus
US-09-734-329-2

Query Match 4.3% Score 229; DB 9; Length 428;
Best Local Similarity 24.0%; Pred. No. 0.00016;
Matches 111; Conservative 26; Mismatches 137; Indels 186; Gaps 20;

OY 130 TAAGGGG---LILASHKL--GATPLPEESTPAPPPPPPPGSGHLNPLILEE 183
| | | | | : | | | | |
Db 37 TTILGKGTKKRYADLSAPKTMGDAPAPFSSINGLSLSPAGSPAP--ASGYAN----- 87
OY 184 LRVLQQRQIHOMQMTQEQICRQVLLLSIGQTVGAPASSELPGTGAASSTKPLPLFSP 243
| | | | | : | | | | |
Db 88 -----DYPFPFSPFGPTGAODPGLV----- 110
OY 244 KPAQOTKTASSSSSSSSSGAEPKQAFHLYHPLGSHQHFV--GGV--GRSHKPTP-- 297
| | | | | : | | | | |
Db 111 -----KXHSDDLPSVYTSLDMTHPYGSWKAGIHAGISPGPGNTPTPMD 157
OY 298 -----APSPALPGSTDL-----IASPHLAFPGT 321
| | | | | : | | | | |
Db 158 MHPRGNMLGGGGGQDGLQGLSTGPAPQPLNPLQPTYSDFAPLNPARYPAHLLQPCP 217
OY 322 TGLL-----AAQCLGARGLEAASPGLLKPKNGSELGCGYEVIS----- 362
| | | | | : | | | | |
Db 218 QHVLPQDVYKPKAVNSGQLGGS--GAAKPPRGAGTGGSGYAGSGRSTDCPCNOC 274
OY 363 LEKPGGR-----HKCRF--CAKVFSGDSALQIHLRSHGERYKCN--VCGNRETT 409
| | | | | : | | | | |
Db 275 LERLGAANAAGLRKKRPHSCHIPDGCYVGAASHLAKHLHMHGGERPVCWMLFCGRKPTR 334
OY 410 RGNLVHFRH-REK-----YPHVQMNHPVP----- 435
| | | | | : | | | | |
Db 335 SDELEHVRHTTRREKFTCLCSKRTFSDHLSKHQRTGEGPGRPPSPKELGEGRSV 394
OY 436 --EHLDYVITSSGLPYGMSVPEKAEEBAGTPGGVKKPLV 475
| | | | | : | | | | |
Db 395 GEEBANOPPRRSTSP---APPEKAH-----GGSPEQGNLL 426

RESULT 31
US-09-864-761-33653
; Sequence 33653, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Acomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263,6
; PRIOR FILING DATE: 2000-10-04
```



```

? PRIOR APPLICATION NUMBER: US_60/236,359
? PRIOR FILING DATE: 2000-09-27
? PRIOR APPLICATION NUMBER: PCT/US01/00666
? PRIOR FILING DATE: 2001-01-30
? PRIOR APPLICATION NUMBER: PCT/US01/00667
? PRIOR FILING DATE: 2001-01-30
? PRIOR APPLICATION NUMBER: PCT/US01/00668
? PRIOR FILING DATE: 2001-01-30
? PRIOR APPLICATION NUMBER: PCT/US01/00669
? PRIOR FILING DATE: 2001-01-30
? PRIOR APPLICATION NUMBER: PCT/US01/00665
? PRIOR FILING DATE: 2001-01-30
? PRIOR APPLICATION NUMBER: PCT/US01/00668
? PRIOR FILING DATE: 2001-01-30
? PRIOR APPLICATION NUMBER: PCT/US01/00663
? PRIOR FILING DATE: 2001-01-30
? PRIOR APPLICATION NUMBER: PCT/US01/00662
? PRIOR FILING DATE: 2001-01-30
? PRIOR APPLICATION NUMBER: PCT/US01/00661
? PRIOR FILING DATE: 2001-01-30
? PRIOR APPLICATION NUMBER: PCT/US01/00670
? PRIOR FILING DATE: 2001-01-30
? PRIOR APPLICATION NUMBER: US_60/234,687
? PRIOR FILING DATE: 2000-09-21
? PRIOR APPLICATION NUMBER: US_09/608,408
? PRIOR FILING DATE: 2000-06-30
? PRIOR APPLICATION NUMBER: US_09/774,203
? PRIOR FILING DATE: 2001-01-29
? NUMBER OF SEQ ID NOS: 49117
? SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
SEQ ID NO 33653
LENGTH: 525
TYPE: prt
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO 284476.6
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 0.8
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.6
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.9
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.2
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.4
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.96
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.2
OTHER INFORMATION: EXPRESSED IN HB.L100, SIGNAL = 2.3
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.5
OTHER INFORMATION: EST_HUMAN HIT: AF150239.1, EVALUOE 1.00e-106
US-09-864-761-33653

Query Match      4.3%   Score 227.5; DB 10; Length 525;
Best Local Similarity 22.5%; Pred.No. 0.00023;
Matches    84; Conservative % 37; Mismatches 134; Indels 119; Gaps 10;

OY       364 EKKGGRIKCKFCFCAVFGSDSALOIHRLRSTGTERPRKYCNWCGNFTTRGNILKVHFRHRREK     423
Db        ||| :||| |::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::|
DB       187 EK*-HECRDCCGAFFTRMLCMHQLIHTGEKPKYCNCGCCAKFOFKNSLTIHGRIHTGE     243

OY       424 YPVMQNPHPVRPHLDIYYTSSGLPEYGMSVPPEKAEEAGTPCGBGVKRPLVAVTALISA     483
Db        |:| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::|
DB       244 KPV-----ECEECG-----KAFSG     257

OY       484 TESLTLLTGTSITAVAPGLPTFNKFVLMAKEPAEKSKADENTPPGSSESAIALVGADSGSAT     543
Db        :|| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::|
DB       258 SSDLT-----KHIRLHTGRPYEGSC-----GRAFSRSSDLSNHK     293

OY       544 RMOLSLIVWSLPBW-----ALLTNHLKSITSFPFYLVLEPIGASBPSETSKILOLYEKI     596
Db        |:| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::|
DB       294 RIHTREKNHYGCPOCGKDFSIKAELTKH-RRIITEERKYRCDECGAKFAHNCKRR-----     346

OY       597 DNGGANAAVASATSPATSAPASSSAGCPNOVCIGLRVLSCLRALRHIGHOGGERPFK     656
Db        -+HERHTGTGKPYPCCROCGKFDXCHCLTLIORHIHTGEKPYK     387

```

```

QY 657 CXYCGRAFSRGNRAHNVGHGHTSPAARAKONSPICQKKTNAVTLDOHVAMHLG----- 711
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 388 CLRCGKAFFSGKSNL-----TNHRIHTGEEKPHKCEVCGMARHSSVILROKHRIHTGEEKPYT 443
QY 712 ----GQIPNGGSAL 721
      |:::|
Db 444 CSECGTSTFRQGSAL 457

RESULT 32
US-09-764-864-956
; Sequence 956, Application US/09764864
; Patent No. US20020132753A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: PT223
CURRENT APPLICATION NUMBER: US/09/764,864
CURRENT FILING DATE: 2001-01-17
Prior application data removed - consult PAM or file wrapper
NUMBER OF SEQ ID NOS: 1792
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 956
LENGTH: 378
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: SITE
LOCATION: (315)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (340)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-864-956

Query Match
Best Local Similarity 22.9%; Score 226; DB 10; Length 378;
Matches 96; Conservative 27; Mismatches 159; Indels 138; Gaps 13;

QY 621 SSASGPNOCVICTLVLSCPRALRLHYGONGGERPKCVCGRAFSRGNRAHNVGHKTS 680
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 10 SHREGYFKCPLCSKVFPPSSLDQHLGDHSSSEHFLCVSCGLFTGTEALLAHRRATPN 69
QY 681 PAARAONSCITCOCKKFNNAVTLQOHVARNHLGGQINRGSGALSGGGAQOINSSQSTASG 740
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 70 PL-----HSCP-CGKTFNLTFLYHRRH-----G 94
QY 741 PGSFPOPOSOOPSEEE-----EMSEEEDEDEEEDVTDSEIAGRS----- 783
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 95 VGVVPLPTTVPV-PEEYVIGFPEPARATGEDEAPPEPYSEESACPARAGTYRCLCSR 153
QY 784 ESGGEKAISVRGDSSEYSGAEVEAVTSVAAPTVEKMSNEKARQNTLPPPPPPDNLDA 843
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 154 EFG--KALQTRHQRFHRLERRHKCSICGKMKKSHVRNHLRTHTGERPFCPPD--- 207
QY 844 PQMEQGTSVSGAMEEBAKLEGISPPAALQLQEGGTSPTLVEELNLRPAMKKDPGESS 903
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 208 -----CSKPRNSPANTLARNHLTHGERP 230
QY 904 GRKACVCGSGSFTOTALLEHQTHPKDGPL-----FT 936
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 231 YR--CGDDCGKAFTQSSSTLRQHLVHAQGFPRKQCGCVAFNRPYRLMNRHYNTGEXPYK 288
QY 937 CYVFCRGFLDRATLKMLLANHVOVPPRAPRQONI-----ATLSLVPG---CSSSTPSP 988
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 289 CRECPRSFTLRLRLLEVHQLVYH-----AGRNHPILPLIMCPLRLTAAPGARLCSRCXQAP 343

RESULT 33
US-09-864-761-42897
; Sequence 42897, Application US/09864761
; Patent No. US20020048763A1

```



```

: GENERAL INFORMATION:
: APPLICANT: Penn, Sharon G.
: APPLICANT: Rank, David R.
: APPLICANT: Hanzel, David K.
: APPLICANT: Chen, Wensheng
: TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
: FILE REFERENCE: Aecm1ca-X-1
: CURRENT FILING DATE: 2001-05-23
: PRIOR APPLICATION NUMBER: US/09/864,761
: PRIOR FILING DATE: 2000-02-04
: PRIOR APPLICATION NUMBER: US 60/180,312
: PRIOR FILING DATE: 2000-02-04
: PRIOR APPLICATION NUMBER: US 60/207,456
: PRIOR FILING DATE: 2000-05-26
: PRIOR APPLICATION NUMBER: US 09/632,366
: PRIOR FILING DATE: 2000-08-03
: PRIOR APPLICATION NUMBER: GB 24263.6
: PRIOR FILING DATE: 2000-10-04
: PRIOR APPLICATION NUMBER: US 60/236,359
: PRIOR FILING DATE: 2000-09-27
: PRIOR APPLICATION NUMBER: PCT/US01/00666
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00667
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00664
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00669
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00665
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00668
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00663
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00662
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00661
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00670
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: US 60/234,687
: PRIOR FILING DATE: 2000-09-21
: PRIOR APPLICATION NUMBER: US 09/608,408
: PRIOR FILING DATE: 2000-06-30
: PRIOR APPLICATION NUMBER: US 09/774,203
: PRIOR FILING DATE: 2001-01-29
: NUMBER OF SEQ ID NOS: 49117
: SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
: SEQ ID NO 42897
: LENGTH: 492
: TYPE: PRT
: ORGANISM: Homo sapiens
: FEATURE:
: OTHER INFORMATION: MAP TO AC007228.1
: OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.1
: OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.1
: OTHER INFORMATION: SWISSPROT HIT: P10078, EVALU0 0.00e+00
: OTHER INFORMATION: EST_HUMAN HIT: AL042491.2, EVALU0 0.00e+00
: OTHER INFORMATION: EST_HUMAN HIT: AL042491.2, EVALU0 0.00e+00
: US-09-864-761-42897

Query Match          4.3%; Score 226; DB 10; Length 492;
Best Local Similarity 22.9%; Pred. No. 0 00027;
Matches 90; Conservative 32; Mismatches 131; Indels 140; Gaps 14;

QY 371 KCRFCAKVGSDSALQIHLRSHTGERPKYKNCVGNRPTTRGNLKVHFRHREKYP----- 425
|||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 121 KCNECKKFTQSSSLVQRHITGKERPKYKNECGKAFSDSSFAHQHQRCHTGKPKYECIE 180
-----HVOQNP-----PYPEHLDYIT 443
QY 426 -----HVOQNP-----PYPEHLDYIT 443
Db 181 CGKAFIQTSLIRHWRYYHTGEKPFDCIDCGKAFSDHIGLNGHRRIRHTGEKPKYKCD--VC 238
```

```

QY 444 SSGLPYGMSVPPKEAEDEAGTP--GGVERKPLVASTALSTESLTJLSTGTSTAVAPGL 502
|||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 239 HKSFRYGGSSSLTVQRHITGKERPKYCDVCRK-----AFSHNLSLT-----QHQRHSGE 286
QY 503 PTEFNKLYLMAVEPKSKADENTPPGSEGSALAGVADSSATRTMQLSKLVTSLPSMALTLN 562
287 KRP-----KCR-----ECGKAFRONIH-----LAS 306
QY 563 HLK-STGSFPFPLYLEPLDGLASPSSETSKL---QOLVEKIDROGAVAVASTAGAPTTAP 618
|||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 307 HLRIHTGEKPEE--CAECGSFSSISQIATHQRI----- 338
QY 619 PSSASGPNQCVICLRVLSQPRALRLHYGHGGERPKCYKAGRAFSTRGNLRAHFVGHK 678
|||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 339 --HTGEKPYECKVCSKAFQTOKAHLAQHOKHTHGEKPYECKGKAFSQ-----TTHLIQHQ 392
QY 679 TSPAARQNSCPICQKKFTTNAVTVLQOHVRMLG 711
|||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 393 RVHTGEKPKYKMECGKAFGDNSSCTOHORLHTG 425
```

RESULT 34

```

US-09-974-298-118
: Sequence 118, Application US/09974298
: Patent No. US20020156263A1
: GENERAL INFORMATION:
: APPLICANT: Chen, Huel-Mei
: TITLE OF INVENTION: GENES EXPRESSED IN BREAST CANCER
: FILE REFERENCE: PA-0037 P
: CURRENT APPLICATION NUMBER: US/09/974,298
: CURRENT FILING DATE: 2001-10-04
: PRIOR APPLICATION NUMBER: 60/238,331
: PRIOR FILING DATE: 2000-05-10
: NUMBER OF SEQ ID NOS: 194
: SOFTWARE: PERL Program
: SEQ ID NO 118
: LENGTH: 516
: TYPE: PRT
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: misc_feature
: OTHER INFORMATION: Incyte ID No. US20020156263A1 5665139CD1
US-09-974-298-118
```

```

Query Match          4.3%; Score 226; DB 9; Length 516;
Best Local Similarity 24.3%; Pred. No. 0.00027;
Matches 87; Conservative 32; Mismatches 133; Indels 106; Gaps 9;

QY 359 VLSLEKPGGRHKRCFCAKVGSDSALQIHLRSHTGERPKYKNCVGNRPTTRGNLKVHFR 418
|||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 253 IHTGEXP---YRCNVCGKRVYHSHLSLAQHQRHITGKERPKYKNECGKAFVSHKSSLVNHR 309
QY 419 RHREKYPHVOQNPVRPEHLDYITSSGLPYGMSVPPKEAEDEAGTPGGVERKPLVAST 478
|||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 310 IHTEKPKYKNECGKAFVSHKSSLVNHRHITGKER--PYKCN-----CGKVFSHSYLA-- 361
QY 479 TALSATSLTLTGTSTAVAPGLPTFNKLYLMAVEPKSKADENTPPGSEGSALAGVAD 538
|||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 362 -----OHLIHAGEKPKYKDECCAKSDNS----- 386
QY 539 SGSATRMQLSKLVTSLPSMALTLNHLKSTGSFPFPLYLEPLDGLASPSSETSKLQ-----QLV 593
|||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 387 -----HLVOHRIHTGE--KRYKDECGKVFSSQNSYLAIVHWRHITG 425
QY 594 EKIDROGAVAVASTAGAPTTAPAPSSASGPNQCVICLRVLSQPRALRLHYGHGGER 653
|||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 426 EK-----AYKNECGKAFGLNLSLAHRRKIHITGKER 455
QY 654 PFKCVGGRATSTRGNLRAHFVGHKTSAPAARQNSCPICQKKFTTNAVTVLQOHVRMLG 711
|||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 456 PFKNECGKAFSSMRSLSLTNHAH--TGEKHFKNECGKLFSDNSYLVLRHORFHAG 509
```


Db 48 FLGAAGAPGSGSSSSSSSSGGGGG-----GGSN-----SSSSSTFN 87
OY LRSHTGERPYKVCNCRRTTRGNLKVHFRHREKYPHVQNMPPHVPREHLDYITSSGLP 448
Db 88 POADTGEQPYE-----HLTAESFPDISLNEKVLVETSYPSQTRLP 129
OY 449 ---YGMVSPPEKAEAEAGT-----PGGVERKPLVASTTALSTATSESL 487
Db 130 PITYTGRSLEAPRNSGNTLWPEPLFSLVSGVSMTPNPASSSSAPSPAASSA-SASOSP 188
OY 488 TLLSTGTST-----AVAGLPTEFNKFLVLMKAVEPKSKADENTPPGSGSAGIAGVADSGS 541
Db 189 PLSCAVPSNDSSPIYSAAPTEPPTNDIF--PEPQSAF---PGSAGTALQYPPPAYP 241
OY 542 ATRMOLSKLVTSIPSWALLTNHLKSTGSPFPYVLEPLGASPSETSKLOQLVEKIDROGA 601
Db 242 AAKGFG--VPMIPDYL-----FPQOGDLGLGTPDOKPFOGL-BSRTQPS 285
OY 602 VAVASTAGAPTTAP-----APSSASGPNQ-----CVI--C 632
Db 286 LTPLTSTIKAFATQSGODIKALNTSYOSOLIKPSRMKRYPNRPSKTPRHERPYACPVESC 345
OY 633 LRVLSCPRALRLHYGOGGERPKVCYGRAFSTRGNLRAHFVGHKTSPARAQNSCPIC 692
Db 346 DRRFSRSDLTFRHRIHTGOKPFCRCIMNFSRSDHLTHIRTH----TGEKPPACDIC 401
OY 693 OKKFTNAVTLQOHVHMLGGQIPNGSALSSEGGAAQENSSSEOSTASGPGSPPOPOSOP 752
Db 402 GRKFARSDERKRHTKIH-----LRQDKKADKSVASSATSSLSYSPS----- 445
OY 753 SPEEMSEEEDEEDEDVTDDESLAGRSESGEKAISVRDSEEVSGAEVATSYA 812
Db 446 -----VAITSYP 451
OY 813 APTTVKEMDSNEKAPQHTLP PPPP-----PDNLDPQPMEDGTSVSGAMEBEAKLEGI 867
Db 452 SVPT-----TSYSPATTSYSPVPTSFSSPSSSTYSPVH-----SGF 490
OY 868 SSPMAALTQEGEGTSPVLEELNLPKMKKDPGSSGRKACEVCGSGFPQTALFEHOKT 927
Db 491 PPSVATTTYS-----SVPPAF---PAOVSSFPSSAVT-NSFSASTGLSDMTAT 534

RESULT 38
US-09-902-941-337
; Sequence 337, Application US/09902941
; Patent No. US20020172952A1
GENERAL INFORMATION:
; APPLICANT: Henderson, Robert A.
; APPLICANT: Wang, Tonglong
; APPLICANT: Watanabe, Yoshitiro
; APPLICANT: Johnson, Jeffrey C.
; APPLICANT: Retter, Mard W.
; APPLICANT: Manerakis, Margarita
; APPLICANT: Carter, Darick
; APPLICANT: Fanger, Gary R.
; APPLICANT: Vedvik, Thomas S.
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: McNabb, Andria
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER
FILE REFERENCE: 210121.478C17
CURRENT APPLICATION NUMBER: US/09/902.941
NUMBER OF SEQ ID NOS: 2002
SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 337
LENGTH: 543
; TYPE: prt
; ORGANISM: Homo sapiens
US-09-902-941-337

Query Match 4.2%; Score 224.5; DB 9; Length 543;

Best Local Similarity 20.6%; Pred. No. 0.00033;
Matches 148; Conservative 67; Mismatches 254; Indels 251; Gaps 28;
OY 269 QAFPHLYHPLGSPHPSVGVGRSHKPPRPAEPALRGSDOLASPHLAFPGTGLLAQ 328
Db 5 KAEMQLMSPLDISDFG---SFPHSPMDNYPKL--BEMMLIS-----NG--APQ 47
OY 329 CLGARGCLEAASPGLIKPNKSGELGVEVLSLEKPGGRHKRCFAKFGSDSALOIH 388
Db 48 FLGAAGAPGSGSSSSSSSGGGGG-----GGSN-----SSSSSTFN 87
OY 389 LRSHTGERPYKVCNCRRTTRGNLKVHFRHREKYPHVQNMPPHVPREHLDYITSSGLP 448
Db 88 POADTGEQPYE-----HLTAESFPDISLNEKVLVETSYPSQTRLP 129
OY 449 ---YGMVSPPEKAEAEAGT-----PGGVERKPLVASTTALSTATSESL 487
Db 130 PITYTGRSLEAPRNSGNTLWPEPLFSLVSGVSMTPNPASSSSAPSPAASSA-SASOSP 188
OY 488 TLLSTGTST-----AVAGLPTEFNKFLVLMKAVEPKSKADENTPPGSGSAGIAGVADSGS 541
Db 189 PLSCAVPSNDSSPIYSAAPTEPPTNDIF--PEPQSAF---PGSAGTALQYPPPAYP 241
OY 542 ATRMOLSKLVTSIPSWALLTNHLKSTGSPFPYVLEPLGASPSETSKLOQLVEKIDROGA 601
Db 242 AAKGFG--VPMIPDYL-----FPQOGDLGLGTPDOKPFOGL-BSRTQPS 285
OY 602 VAVASTAGAPTTAP-----APSSASGPNQ-----CVI--C 632
Db 286 LTPLTSTIKAFATQSGODIKALNTSYOSOLIKPSRMKRYPNRPSKTPRHERPYACPVESC 345
OY 633 LRVLSCPRALRLHYGOGGERPKVCYGRAFSTRGNLRAHFVGHKTSPARAQNSCPIC 692
Db 346 DRRFSRSDLTFRHRIHTGOKPFCRCIMNFSRSDHLTHIRTH----TGEKPPACDIC 401
OY 693 OKKFTNAVTLQOHVHMLGGQIPNGSALSSEGGAAQENSSSEOSTASGPGSPPOPOSOP 752
Db 402 GRKFARSDERKRHTKIH-----LRQDKKADKSVASSATSSLSYSPS----- 445
OY 753 SPEEMSEEEDEEDEDVTDDESLAGRSESGEKAISVRDSEEVSGAEVATSYA 812
Db 446 -----VAITSYP 451
OY 813 APTTVKEMDSNEKAPQHTLP PPPP-----PDNLDPQPMEDGTSVSGAMEBEAKLEGI 867
Db 452 SVPT-----TSYSPATTSYSPVPTSFSSPSSSTYSPVH-----SGF 490
OY 868 SSPMAALTQEGEGTSPVLEELNLPKMKKDPGSSGRKACEVCGSGFPQTALFEHOKT 927
Db 491 PPSVATTTYS-----SVPPAF---PAOVSSFPSSAVT-NSFSASTGLSDMTAT 534

RESULT 39
US-09-849-626-337
; Sequence 337, Application US/09849626
; Publication No. US20020197669A1
GENERAL INFORMATION:
; APPLICANT: Bangur, Chaitanya
; APPLICANT: Fanger, Gary
; APPLICANT: Wang, Tonglong
; APPLICANT: Switzer, Anne
; APPLICANT: McNeill, Patricia
; APPLICANT: Clapper, Jonathan
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF LUNG CANCER
FILE REFERENCE: 210121.478C16
CURRENT APPLICATION NUMBER: US/09/849.626
NUMBER OF SEQ ID NOS: 1926
SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 337
LENGTH: 543

TYPE: PRT
ORGANISM: Homo sapiens
US-09-949-626-337

Query Match 4.2%; Score 224.5; DB 9; Length 543;

Best Local Similarity 20.6%; Pred. No. 0.00033;

Matches 148; Conservative 67; Mismatches 25; Indels 25; Gaps 28;

```
269 QAFHLHPLGSGHPESVGVGRSHKPTAPSPALPGSTDOIILASPHLAFPGTTGLLAQ 328
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 5 KAEMQLMSPLQISDPG----SFPHSPTMDNYPKL---EEMLLS-----NG---APQ 47

329 CUGARGLEAASPGLLKPKNGSELGVEYISLEKGGRRKRCRCACAVFGSDSLQIH 388
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 48 FLGAAGAPRGSGNSSSSSSGGGGGG-----GGSN-----SSSSSTFN 87

389 LSHNGERPCKVCNCRFTTNGMLKVHHRREKYPHYQMPPHYPEHLDYITSSGLP 448
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 88 PPDNGEOPYE-----HITAESFPDISLNNKVLVETSYPSOTRLP 129

449 ---YGNVPEPKAEERAGT-----PEGVERKPLVASTALSLATESL 487
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 130 PLYTGREFSLEPAPNSGNTLMPEDFLSVGLVSMTNPPASSSSAPSPAASSA--SASQSP 188

488 TLLSTGTST-----AVAPGLPTFNKFLMKAVEBKSKADENTPPSEGSALGVADSGS 541
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 189 PLSCAVPSNDSPSYSAAPFTPTPTDIF--PEPOSQAF---PESAGTALQPPPAYP 241

542 ATPMQLSKLVTSLPSWALLTNHLKSTGSPFPYVLEPLGASPSSETSKLOLVEKIDROGA 601
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 242 AAKGGQ--VPMIDPYL-----FPQGGDGLGTFPDQKPFQGL--ESRTQOPS 285

602 VAVASTASGAPTSAP-----APSSASGPNQ-----CVI--C 632
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 286 LPLSTRIKATATGSGSODLKALNTSYOSQLIKPSRMKYPNRPSPKTPHPRYPACPVESC 345

633 LRVLSCELRALRHVGOGGERPFKCVGGRFSTRGNLRAHFHGHTKSPARAQNSCPIC 692
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 346 DRPRSRDELRRHRIHGOXKPFQCRICKRNFSRSDHLTHITH----TGERKPADIC 401

693 QKRTTNVTLQOAHYRMLGQIIPNGSALSEGGAQAQNSSEOSTASGPGSFPPOSQOP 752
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 402 GRFARSDERKRRHKIH-----LRQDKRADKSVASATSSLSYPSP----- 445

753 SPREEMGDEEEDDEEDVDDESLAGSGESGGKALSVRQDSEVSAGAEVATYVA 812
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 446 -----VAISYP 451

813 APPTVKKEDSNEKAPOHTLPPPPP-----PNDLHPQPMEGCTDVSGAMEERAKLEG 867
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 452 SPVT--TSYSPATTSTYSPVPTSPSSPGSSTYPSPVH-----SGF 490

868 SSPMAALQOEGETSTPLVEELNLPKAMKDPGESSGRKACEVCGSGFPOTALEEHOKT 927
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 491 PSEPVAATYTS-----SVPPAF--PAOVSSFPSSAVT--NSFASASTGLSDMTAT 534
```

RESULT 40

US-09-929-315-4

Sequence 4, Application US/09929315

Patent No. US20020082394A1

GENERAL INFORMATION:

APPLICANT: Call, Katherine M.
APPLICANT: Glaser, Thomas M.
APPLICANT: Ito, Caryn Y.
APPLICANT: Buckler, Alan J.
APPLICANT: Pelletier, Jerry
APPLICANT: Haber, Daniel A.
APPLICANT: Rose, Elise A.
APPLICANT: Housman, David E.
APPLICANT: Bruening, Wendy
APPLICANT: Darveau, Andre
TITLE OF INVENTION: Localization and Characterization of the

```
;; TITLE OF INVENTION: Wilms' Tumor Gene
;; FILE REFERENCE: 0050.1312-013
;; CURRENT APPLICATION NUMBER: US/09/929,315
;; CURRENT FILING DATE: 2001-08-14
;; PRIOR APPLICATION NUMBER: US 09/037,179
;; PRIOR FILING DATE: 1998-03-09
;; PRIOR APPLICATION NUMBER: US 08/102,942
;; PRIOR FILING DATE: 1993-08-02
;; PRIOR APPLICATION NUMBER: US 07/614,161
;; PRIOR FILING DATE: 1990-11-13
;; PRIOR APPLICATION NUMBER: US 07/435,780
;; PRIOR FILING DATE: 1989-11-13
;; PRIOR APPLICATION NUMBER: US 07/795,323
;; PRIOR FILING DATE: 1991-11-20
;; NUMBER OF SEQ ID NOS: 21
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 4
;; LENGTH: 449
;; TYPE: PRT
;; ORGANISM: Unknown
;; FEATURE:
;; OTHER INFORMATION: Murine
;; US-09-929-315-4
```

Query Match 4.2%; Score 222; DB 10; Length 449;

Best Local Similarity 25.6%; Pred. No. 0.00036; Matches 100; Conservative 38; Mismatches 135; Indels 118; Gaps 20;

```
QY 124 LVNATGTAAGGGGLILASPKLGA---TPL-----PEST-----PAPPPPPPPPPG 169
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 11 LPLAVSSILGGGGGCGL--PVSGARQMAFVLDPAFPGASVYGLGFAFPAPPPPPPP 67

170 VSGCHLNPILLEELRYLQORIHOMQMTQICROYLLLSLQO--TVGA-----PA 219
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 68 -----PHSFIQDEPSWGAEPHEQCLSAFTLHFSQFTGTAGACRYGPPGP 115

220 SPSEL-PEFGAASSTKPLPLFSPIKPA--QTGKTTA-----SSSSSSSSGAEPKQA 270
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 116 PPSQASSGQAMFPAFVPLPSCLESQPTIRNOGYSTVFEDGAPSYGHTPSHAAQFPNHS 175

271 FFIHLHPLGSGHPESVGVGRSHKPTP-----ASPALPGSTDOIILASPHLA---FPG 320
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 176 FKH-EDPQGQO-----GSLGEQOQSYVPPPYGCHTPTDSCGTSALLLRPYSSDNLXQM 229

321 TTGLLAQC-----LGAARGLAASPGLLKPKNGSGELGYEYISLEKP--GGRH 370
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 230 TSOL--ECMTWNNQNLGATLTKGMAAGSSSVKWTGQSNHGTGESENHHTAPILCGAOY 286

QY 371 K-----CRF--CAKYFGSDSLQIHLRS 391
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 287 RIHTHGVRGIQDVVRVSGVAPTLVRSASSETSEKRPFMCAYPGCNKRYFKLSHLQMSRK 346

QY 392 HTGERPYKCNV--CGNRFTRGNLKVHFRH 420
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 347 HTGERPYQCDFRKDERFRSRSDQLKRHQRH 377
```

Search completed: January 13, 2003, 15:23:42
Job time : 31.9851 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 13, 2003, 15:17:40 : Search time 38.0568 Seconds
(without alignments)
3518.868 Million cell updates/sec

Title: US-09-988-117-1
Perfect score: 5277
Sequence: IMAHERSSSRKGVAPAGEPAE.....PITSTGLSPFPKKDPTIP 1005

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues
Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 08
Maximum Match 1008
Listing first 100 summaries

Database :

_A_Geneseq_101002:*

1: /SIDs2/gcgcdata/geneseq/geneseq-emb1/AA1980.DAT:*
2: /SIDs2/gcgcdata/geneseq/geneseq-emb1/AA1981.DAT:*
3: /SIDs2/gcgcdata/geneseq/geneseq-emb1/AA1982.DAT:*
4: /SIDs2/gcgcdata/geneseq/geneseq-emb1/AA1983.DAT:*
5: /SIDs2/gcgcdata/geneseq/geneseq-emb1/AA1984.DAT:*
6: /SIDs2/gcgcdata/geneseq/geneseq-emb1/AA1985.DAT:*
7: /SIDs2/gcgcdata/geneseq/geneseq-emb1/AA1986.DAT:*
8: /SIDs2/gcgcdata/geneseq/geneseq-emb1/AA1987.DAT:*
9: /SIDs2/gcgcdata/geneseq/geneseq-emb1/AA1988.DAT:*
10: /SIDs2/gcgcdata/geneseq/geneseq-emb1/AA1989.DAT:*
11: /SIDs2/gcgcdata/geneseq/geneseq-emb1/AA1990.DAT:*
12: /SIDs2/gcgcdata/geneseq/geneseq-emb1/AA1991.DAT:*
13: /SIDs2/gcgcdata/geneseq/geneseq-emb1/AA1992.DAT:*
14: /SIDs2/gcgcdata/geneseq/geneseq-emb1/AA1993.DAT:*
15: /SIDs2/gcgcdata/geneseq/geneseq-emb1/AA1994.DAT:*
16: /SIDs2/gcgcdata/geneseq/geneseq-emb1/AA1995.DAT:*
17: /SIDs2/gcgcdata/geneseq/geneseq-emb1/AA1997.DAT:*
18: /SIDs2/gcgcdata/geneseq/geneseq-emb1/AA1998.DAT:*
19: /SIDs2/gcgcdata/geneseq/geneseq-emb1/AA1999.DAT:*
20: /SIDs2/gcgcdata/geneseq/geneseq-emb1/AA2000.DAT:*
21: /SIDs2/gcgcdata/geneseq/geneseq-emb1/AA2001.DAT:*
22: /SIDs2/gcgcdata/geneseq/geneseq-emb1/AA2002.DAT:*
23: /SIDs2/gcgcdata/geneseq/geneseq-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysts of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	5277	100.0	1005	23	AAE17954 Human Sal2 protein
2	5272	99.9	1005	23	AAE17967 Human Sal2 protein
3	5269	99.8	1005	23	AAE17968 Human Sal2 protein
4	5171	98.0	1007	22	AAW78838 Human protein SEQ
5	5161	97.8	1019	22	AAW78822 Human protein SEQ
6	4677.5	88.6	1002	23	AAE17955 Mouse Sal2 protein
7	877.5	16.6	813	22	AAE93193 Human protein sequ
8	784	14.9	1373	22	ABE62368 Drosophila melanog
9	733	13.9	172	22	AAO01962 Human polypeptide
10	726	13.8	1267	22	ABE61050 Drosophila melanog

11	596	11.3	200	22	AAE92796 Human protein sequ
12	489	9.3	330	22	ABE30894 Peptide #3545 enco
13	489	9.3	330	22	ABE36076 Peptide #3582 enco
14	489	9.3	330	22	AAE56858 Human brain expres
15	489	9.3	330	22	AAE69245 Human bone marrow
16	489	9.3	330	22	AAE17074 Peptide #3508 enco
17	489	9.3	330	22	AAE29568 Peptide #3605 enco
18	489	9.3	330	22	AAE04775 Peptide #3457 enco
19	489	9.3	330	22	ABE38855 Human peptide enco
20	441	8.4	336	22	ABE32272 Peptide #5036 enco
21	441	8.4	336	22	ABE37530 Human brain expres
22	441	8.4	336	22	AAE58186 Human bone marrow
23	441	8.4	336	22	AAE70641 Peptide #4919 enco
24	441	8.4	336	22	AAE18485 Peptide #4993 enco
25	441	8.4	336	22	AAE30956 Peptide #4757 enco
26	441	8.4	336	22	AAE06075 Human peptide enco
27	441	8.4	336	22	ABE40328 Human zinc finger
28	343	6.5	702	23	ABE04333 Human nucleic acid
29	341	6.5	799	21	AAE21003 Human protein SEQ
30	334	6.3	927	22	AAE80283 Human protein sequ
31	333.5	6.3	619	22	AAE95103 Human protein sequ
32	329.5	6.2	498	22	AAE94669 Human protein sequ
33	324.5	6.1	725	22	AAE95116 Human protein SEQ
34	320.5	6.1	869	22	AAE79299 Novel human diagno
35	311.5	5.9	882	22	ABE06465 Drosophila melanog
36	311.5	5.9	934	22	ABE69958 Drosophila melanog
37	308.5	5.8	1893	22	ABE59829 Human polypeptide
38	303	5.7	722	22	AAE39029 Drosophila melanog
39	298.5	5.7	744	22	ABE59449 Drosophila melanog
40	298.5	5.7	744	22	ABE66002 Drosophila melanog
41	298.5	5.7	744	22	ABE65712 Drosophila melanog
42	295	5.6	496	22	AAE95788 Human protein sequ
43	293.5	5.6	1212	22	ABE00399 Novel human diagno
44	293.5	5.6	1472	22	AAE31578 Drosophila melanog
45	293	5.6	820	22	ABE55565 Mouse combined DNA
46	293	5.6	884	22	AAE99334 Novel human diagno
47	291	5.5	1685	22	ABE62718 Novel human diagno
48	290.5	5.5	691	22	AAE94433 Human atrophin I p
49	289.5	5.5	1185	20	AAE33497 Mouse Rlt1 alpha'
50	289	5.5	880	22	AAE99336 Novel human secret
51	288	5.5	1220	22	AAE10831 Novel human diagno
52	285.5	5.4	832	22	ABE10543 Novel human diagno
53	284.5	5.4	1156	22	ABE628743 Novel human diagno
54	284	5.4	582	23	ABE97429 Novel human diagno
55	283.5	5.4	883	22	ABE62743 Novel human diagno
56	283.5	5.4	1252	22	AAE79739 Human protein SEQ
57	282	5.3	613	22	AAE95862 Human secreted pro
58	281.5	5.3	451	21	AAE52155 Human secreted co
59	281.5	5.3	624	23	ABE57324 Human Rlt1 protein
60	281.5	5.3	894	22	AAE89335 Novel human diagno
61	280.5	5.3	719	22	ABE16953 Novel human diagno
62	279.5	5.3	973	22	ABE21022 Novel human diagno
63	278.5	5.3	890	22	AAE99339 Human Rlt1 alpha'
64	278	5.3	755	22	AAE40916 Human polypeptide
65	277	5.2	1054	22	ABE57993 Drosophila melanog
66	276.5	5.2	742	23	ABE05703 Human nucleic acid
67	276.5	5.2	1342	22	ABE01726 Novel human diagno
68	275	5.2	809	22	AAE38689 Human polypeptide
69	275	5.2	2703	22	ABE63299 Drosophila melanog
70	274.5	5.2	730	22	ABE93278 Human protein sequ
71	273.5	5.2	356	22	ABE60575 Drosophila melanog
72	273	5.2	675	22	AAE94388 Human protein sequ
73	273	5.2	714	22	ABE17368 Novel human diagno
74	273	5.2	812	22	AAE99337 Mouse Rlt1 beta pr
75	272.5	5.2	878	22	AAE29528 Human protein sequ
76	271.5	5.1	803	22	AAE95278 Human protein sequ
77	271	5.1	823	22	AAE99340 Human Rlt1 beta pr
78	270	5.1	839	22	AAE95280 Human protein sequ
79	269.5	5.1	688	23	AAE9480 Human zinc finger
80	269.5	5.1	720	22	AAE94329 Human protein sequ
81	269.5	5.1	1203	22	ABE61375 Drosophila melanog
82	269	5.1	751	22	AAE39508 Human polypeptide
83	268.5	5.1	841	22	ABE16167 Novel human diagno

84	268	5.1	567	22	AAM38658	Human polypeptide
85	268	5.1	751	22	AAM39130	Human polypeptide
86	268	5.1	1205	22	ABG05068	Novel human diago
87	268	5.1	1205	22	ABG15312	Novel human diago
88	268	5.1	1214	22	ABG27121	Novel human diago
89	267.5	5.1	515	22	ABB31309	Peptide #3960 enco
90	267.5	5.1	515	22	ABB36515	Peptide #4021 enco
91	267.5	5.1	515	22	ABB21854	Protein #3853 enco
92	267.5	5.1	515	22	AAM57278	Human brain expres
93	267.5	5.1	515	22	AAM69683	Human bone marrow
94	267.5	5.1	515	22	AAM17493	Peptide #3929 enco
95	267.5	5.1	515	22	AAM30016	Peptide #4053 enco
96	267.5	5.1	515	22	AAM05170	Peptide #3852 enco
97	267.5	5.1	515	23	ABG39299	Human peptide enco
98	266.5	5.1	2639	22	ABG15016	Novel human diago
99	266	5.0	577	22	AAU15930	Human novel secret
100	266	5.0	645	22	ABG28285	Novel human diago

ALIGNMENTS

RESULT 1

ID AAE17954 standard; protein; 1005 AA.

XX AAE17954;

DT 07-MAY-2002 (first entry)

XX Human Sal2 protein.

KW Human; proliferative disorder; tumor host range mutant virus; cancer;

KV T-HR mutant; Sal2 protein; ovarian tumour; chromosome 14q12.

XX Homo sapiens.

XX Key Location/Qualifiers

FT Misc-difference 15

FT /note= "Encoded by TGC of the inverse complementary strand of the sequence shown as SEQ ID NO:2 (AAD28651)"

FT Misc-difference 22..23

FT /note= "Encoded by bases 4611-5022 of the inverse complementary strand of the sequence shown as

FT Misc-difference 120

FT /note= "Encoded by CCA of the inverse complementary strand of the sequence shown as SEQ ID NO:2 (AAD28651)"

FT Misc-difference 545

FT /note= "Encoded by CGC of the inverse complementary strand of the sequence shown as SEQ ID NO:2 (AAD28651)"

FT Misc-difference 552

FT /note= "Encoded by GTG of the inverse complementary strand of the sequence shown as SEQ ID NO:2 (AAD28651)"

FT Misc-difference 573

FT /note= "Encoded by TTC of the inverse complementary strand of the sequence shown as SEQ ID NO:2 (AAD28651)"

FT Misc-difference 575

FT /note= "Encoded by TAT of the inverse complementary strand of the sequence shown as SEQ ID NO:2 (AAD28651)"

FT Misc-difference 576

FT /note= "Encoded by GTG of the inverse complementary strand of the sequence shown as SEQ ID NO:2 (AAD28651)"

FT Misc-difference 577

FT /note= "Encoded by ATC of the inverse complementary strand of the sequence shown as SEQ ID NO:2 (AAD28651)"

FT Misc-difference 578

FT /note= "Encoded by GAG of the inverse complementary strand of the sequence shown as SEQ ID NO:2 (AAD28651)"

FT Misc-difference 579

FT /note= "Encoded by CCC of the inverse complementary strand of the sequence shown as SEQ ID NO:2 (AAD28651)"

PN	NO200204596-A2.
XX	
XX	17-JAN-2002.
PD	
XX	
XX	05-JUL-2001; 2001WO-US21354.
PF	
XX	
PR	07-JUL-2000; 2000US-216723P.
PR	19-MAR-2001; 2001US-0812471.
PR	19-MAR-2001; 2001US-0812633.
XX	
XX	(HARD) HARVARD COLLEGE.
PA	(BGHM) BRIGHAM & WOMENS HOSPITAL INC.
PI	Benjamin TL, Li D, Mok SC, Cramer DW, Ma Y;
XX	
XX	WPI; 2002-164637/21.
DR	N-PSDB; AAD28651.
DR	
XX	
PT	Detecting protein involved in susceptibility to proliferative disease,
PT	by infecting normal and abnormal proliferating cells with mutant virus,
PT	- detecting mutated protein allowing growth of mutant on abnormal cells
PT	
XX	
XX	
PS	Disclosure; Page 77-79; 92pp; English.
XX	
XX	

The invention relates to a method for the identification of genes and their encoded proteins involved in susceptibility to proliferative disorders, including cancer using a tumor host range mutant (T-HR mutant) virus. The invention also provides the use of Sal2 genes and proteins in methods of identifying a mammal having, or at a risk of acquiring a proliferative disease. T-HR mutants are used to kill cancer cells such as one carrying a Sal2 alteration. Transgenic and knockout mouse comprising Sal2 nucleic acid are useful as research tools to determine genetic and physiological features of cancer and for identifying compounds that can affect ovarian and other tumours. The present sequence is human Sal2 protein. The Sal2 gene is located on chromosome 14q12.

Sequence 1005 AA:

Query Match 100.0%; Score 5277; DB 23; Length 1005;
Best Local Similarity 100.0%; Pred. No. 4.2e-295;
Matches 1005; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MAHESRSSRLGVAPGPAETGSDASEDDHPOVCAKCAQFTDPTEFLAHQNAKSTDPV	60
DB	1	MAHESRSSRLGVAPGPAETGSDASEDDHPOVCAKCAQFTDPTEFLAHQNAKSTDPV	60
QY	61	MYLIGQENPNSSASSEPREGHNPPQVMDTEHSPDPSSGVPTDPTWGPERRGEES	120
DB	61	MYLIGQENPNSSASSEPREGHNPPQVMDTEHSPDPSSGVPTDPTWGPERRGEES	120
QY	121	GHTLVATGTGAAGGGGLILASPKLGATPLPEESTPAPPPPPPPPPGVGSHLNIPLT	180
DB	121	GHTLVATGTGAAGGGGLILASPKLGATPLPEESTPAPPPPPPPPPGVGSHLNIPLT	180
QY	181	LEELRLVLOQROIHQOMTTEOICROVLLLSIGOTGAPASPELDTGTASSKPLPLPT	240
DB	181	LEELRLVLOQROIHQOMTTEOICROVLLLSIGOTGAPASPELDTGTASSKPLPLPT	240
QY	241	SPKPVQTSKTLAASSSSSSSGAETPKQAFPHLYHPLGSHPPSAGVGSHKPTPAP	300
DB	241	SPKPVQTSKTLAASSSSSSSGAETPKQAFPHLYHPLGSHPPSAGVGSHKPTPAP	300
QY	301	SPALPESTQOLILASPHLAFSTTGLLAAQCI GAARGLAATASPGLIKRNGSGELSTGEV	360
DB	301	SPALPESTQOLILASPHLAFSTTGLLAAQCI GAARGLAATASPGLIKRNGSGELSTGEV	360
QY	361	MGPLEKPGGRHKCRFCAKYVGSALQIHLRSHTGERPKCNVGNRPPTTGNLKVHFR	420
DB	361	MGPLEKPGGRHKCRFCAKYVGSALQIHLRSHTGERPKCNVGNRPPTTGNLKVHFR	420
QY	421	HREKTPHVOGNPDPVPEHLDTYITSSGLPYGMSVPEKAEDEAATPGGVERKPLVASTT	480


```
Db 421 HREKYPHQMNPHPVEHLDYITSSGLPRGMSVPPKAEAEATPGGVERKPLVASTT 480
Qy 481 ALSATESLTLSTAGTATAPGLPAFNKFLMKAVEPKNADENTPGSGSAISGVAES 540
Db 481 ALSATESLTLSTAGTATAPGLPAFNKFLMKAVEPKNADENTPGSGSAISGVAES 540
Qy 541 STATLMQLSKLTSLPSMALLTNHFKSTGSPFLPLCARALGASPSSTKLOQLVEKIDRQ 600
Db 541 STATLMQLSKLTSLPSMALLTNHFKSTGSPFLPLCARALGASPSSTKLOQLVEKIDRQ 600
Qy 601 GAVAVTSAASGAPTTSAFAPSSASSGPNOCVLCRLVLSCPRALRLHYGQHGGERPFCK 660
Db 601 GAVAVTSAASGAPTTSAFAPSSASSGPNOCVLCRLVLSCPRALRLHYGQHGGERPFCK 660
Qy 661 VCGRAESTRGNLRAHFVGHKASPAARAQNSCPICQKKFTNAVTLOQHYRMHLGGQIPNGG 720
Db 661 VCGRAESTRGNLRAHFVGHKASPAARAQNSCPICQKKFTNAVTLOQHYRMHLGGQIPNGG 720
Qy 721 TALPBGGAAGNGSFQSTVSGAGSPFOQSQSPSEELSEEEDEDEEDVDYDEDS 780
Db 721 TALPBGGAAGNGSFQSTVSGAGSPFOQSQSPSEELSEEEDEDEEDVDYDEDS 780
Qy 781 LAGRSESGGEKAISVRGDSSEASGAEEVGTVAATAATAGKEMDSNEKTTQOSSLPPEPP 840
Db 781 LAGRSESGGEKAISVRGDSSEASGAEEVGTVAATAATAGKEMDSNEKTTQOSSLPPEPP 840
Qy 841 PDSLDPQPMEOGSSGVLGKEGCKPERSSSPASALTPEGEATSVTLVEELSLQEMARK 900
Db 841 PDSLDPQPMEOGSSGVLGKEGCKPERSSSPASALTPEGEATSVTLVEELSLQEMARK 900
Qy 901 EPGESSSRKACEVCGAFPSQALIEHQKTPKRGPLTCVFCRQGFLEKATLKHMLLA 960
Db 901 EPGESSSRKACEVCGAFPSQALIEHQKTPKRGPLTCVFCRQGFLEKATLKHMLLA 960
Qy 961 HHQVOPFAPHGPNIALSLVPGCSPTSITSTGLSPFPKKDDPTIP 1005
Db 961 HHQVOPFAPHGPNIALSLVPGCSPTSITSTGLSPFPKKDDPTIP 1005

RESULT 2
AAE17967 standard; Protein: 1005 AA.
ID AAE17967
AC AAE17967;
DT 07-MAY-2002 (first entry)
DE Human Sal2 protein mutant (S73C).
KW Human: proliferative disorder; tumor host range mutant virus; cancer;
T-HR mutant; Sal2 protein; ovarian tumour; mutant; mutcin.
XX Homo sapiens.
OS Synthetic.
XX Key Location/Qualifiers
FH Misc-difference 73 /note= "Wild-type Ser substituted with Cys"
FT WT
XX WO200204596-A2.
XX 17-JAN-2002.
XX 05-JUL-2001: 2001MO-US21354.
XX 07-JUL-2000: 2000US-216723P.
XX 19-MAR-2001: 2001US-0812471.
XX 19-MAR-2001: 2001US-0812633.
XX (HARD ) HARVARD COLLEGE.
XX (BGHM ) BRIGHAM & WOMENS HOSPITAL INC.
```

```
PI Benjamin TL, Li D, Mok SC, Cramer DW, Ma Y;
XX WPL: 2002-164637/21.
DR
XX
XX Detecting protein involved in susceptibility to proliferative disease,
PT by infecting normal and abnormal proliferating cells with mutant virus,
PT detecting mutated protein allowing growth of mutant on abnormal cells
PT
XX
XX Claim 35; Page -: 92pp; English.
XX
XX The invention relates to a method for the identification of genes
CC and their encoded proteins involved in susceptibility to proliferative
CC disorders, including cancer using a tumor host range mutant (T-HR
CC mutant) virus. The invention also provides the use of Sal2 genes and
CC mutants in methods of identifying a mammal having, or at a risk of
CC acquiring a proliferative disease. T-HR mutants are used to kill cancer
CC cells such as one carrying a Sal2 alteration. Transgenic and knockout
CC mouse comprising Sal2 nucleic acid are useful as research tools to
CC determine genetic and physiological features of cancer and for
CC identifying compounds that can affect ovarian and other tumours. The
CC present sequence is human Sal2 protein mutant (S73C).
CC Note: This sequence is not shown in the specification but is derived
CC from the human Sal2 wild-type protein shown as SEQ ID NO:1 (AAE17964)
CC in page 77-79 of the specification.
XX
XX Sequence 1005 AA:
SQ
XX
XX Query Match 99.9%; Score 5272; DB 23; Length 1005;
XX Best Local Similarity 99.9%; Pred. No. 8,2e-295;
XX Matches 1004; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 MAHERSSRLGYPAGEPAELGDSBEDHPYCKCAQFTDPTFEFLAHQNACTDPPV 60
Db 1 MAHERSSRLGYPAGEPAELGDSBEDHPYCKCAQFTDPTFEFLAHQNACTDPPV 60
Qy 61 MYIIGQENPNNSASSEPREGHNNPQVMDTEHSNPDGSSVPTDPTWGPERRGEES 120
Db 61 MYIIGQENPNNSASSEPREGHNNPQVMDTEHSNPDGSSVPTDPTWGPERRGEES 120
Qy 121 GHRVATATGTAAGGGGLITASPRLGATPLPESTPAPPPPPPPPGVSGHNLNPLI 180
Db 121 GHRVATATGTAAGGGGLITASPRLGATPLPESTPAPPPPPPPPGVSGHNLNPLI 180
Qy 181 LEEELRVLOQRQIHQMOMTEQICRQVLLGSLQGTGAPAPSPSELPCTGTAASSTRPLPLF 240
Db 181 LEEELRVLOQRQIHQMOMTEQICRQVLLGSLQGTGAPAPSPSELPCTGTAASSTRPLPLF 240
Qy 241 SPIKPVQTSKTLASSSSSSSSGAETPKQAFPHLYHPLGSOHPFSAGVGRSHKPTPAP 300
Db 241 SPIKPVQTSKTLASSSSSSSSGAETPKQAFPHLYHPLGSOHPFSAGVGRSHKPTPAP 300
Qy 301 SPALPGSTQDLIASPHLAFSTTGTLAOCIGAARGLEATPAGGLKPKNGSGELSYGEV 360
Db 301 SPALPGSTQDLIASPHLAFSTTGTLAOCIGAARGLEATPAGGLKPKNGSGELSYGEV 360
Qy 361 MCPLEKPGRHKCFKCAKVFSDSALQIHLRSHTGERPYKCNVCGNFTTNGNLKVHNR 420
Db 361 MCPLEKPGRHKCFKCAKVFSDSALQIHLRSHTGERPYKCNVCGNFTTNGNLKVHNR 420
Qy 421 HREKYPHQMNPHPVEHLDYITSSGLPRGMSVPPKAEAEATPGGVERKPLVASTT 480
Db 421 HREKYPHQMNPHPVEHLDYITSSGLPRGMSVPPKAEAEATPGGVERKPLVASTT 480
Qy 481 ALSATESLTLSTAGTATAPGLPAFNKFLMKAVEPKNADENTPGSGSAISGVAES 540
Db 481 ALSATESLTLSTAGTATAPGLPAFNKFLMKAVEPKNADENTPGSGSAISGVAES 540
Qy 541 STATLMQLSKLTSLPSMALLTNHFKSTGSPFLPLCARALGASPSSTKLOQLVEKIDRQ 600
Db 541 STATLMQLSKLTSLPSMALLTNHFKSTGSPFLPLCARALGASPSSTKLOQLVEKIDRQ 600
Qy 601 GAVAVTSAASGAPTTSAFAPSSASSGPNOCVLCRLVLSCPRALRLHYGQHGGERPFCK 660
```


|||||
Db GAVAVTSAASGAPTTSAAPASSSSAGSPNCVICTLRVLSCPRALRLHYGQHGGERPFCK 660
Qy 661 VCGRAFSTRGNLRAHFVGHKASPAARAONSCPICOKKFTNAVTLQOHVRLHMLGGQIPNGG 720
Db 661 VCGRAFSTRGNLRAHFVGHKASPAARAONSCPICOKKFTNAVTLQOHVRLHMLGGQIPNGG 720
Qy 721 TALEGGGAOAENGSEOSTVSGAGSPPOOQSOQSPPEELSEEEDEDEEDVDDED 780
Db 721 TALEGGGAOAENGSEOSTVSGAGSPPOOQSOQSPPEELSEEEDEDEEDVDDED 780
Qy 781 LAGRSGSGGKALISVRGDSEBASGAEEVGTVAATAATAGKEMDNKKTQOSSLPPPP 840
Db 781 LAGRSGSGGKALISVRGDSEBASGAEEVGTVAATAATAGKEMDNKKTQOSSLPPPP 840
Qy 841 PDSLDOPOMPOGSSGVGKREGKPERSSSPASALTPEGEATSVTLVEELSLQEAARK 900
Db 841 PDSLDOPOMPOGSSGVGKREGKPERSSSPASALTPEGEATSVTLVEELSLQEAARK 900
Qy 901 EPGSSSRKACEVGOAPPSQALIEHOKTHPEKPLETCVFCROGFLERATLKKHMLA 960
Db 901 EPGSSSRKACEVGOAPPSQALIEHOKTHPEKPLETCVFCROGFLERATLKKHMLA 960
Qy 961 HHQVPPAPHPQONIALSLVPGSPSTSTGSLSPPRKDDPTIP 1005
Db 961 HHQVPPAPHPQONIALSLVPGSPSTSTGSLSPPRKDDPTIP 1005
RESULT 3
AAE17968
ID AAE17968 standard; Protein: 1005 AA.
AC AAE17968;
DT 07-MAY-2002 (first entry)
DE Human Sal2 protein mutant (G744R).
KW Human; proliferative disorder; tumor host range mutant virus; cancer;
T-HR mutant; Sal2 protein; ovarian tumour; mutant; muteln.
OS Homo sapiens.
XX Synthetic.
FH Key Location/Qualifiers
FT Misc-difference 744 /note= "Wild-type Gly substituted with Arg"
XX PN MO200204596-A2.
XX PD 17-JAN-2002.
XX PE 05-JUL-2001; 2001MO-US21354.
XX PR 07-JUL-2000; 2000US-216723P.
XX PR 19-MAR-2001; 2001US-0812471.
XX PR 19-MAR-2001; 2001US-0812633.
XX PA (HARD) HARVARD COLLEGE.
XX PA (BGHM) BRIGHAM & WOMENS HOSPITAL INC.
XX PI Benjamin TL, Li D, Mok SC, Cramer DW, Ma Y;
XX WPI: 2002-164637/21.
XX DR
XX PT Detecting protein involved in susceptibility to proliferative disease,
XX PT by infecting normal and abnormal proliferating cells with mutant virus,
XX PT detecting mutated protein allowing growth of mutant on abnormal cells
XX PT
XX Example 7; Page -: 92pp; English.
XX PS
XX The invention relates to a method for the identification of genes

CC and their encoded proteins involved in susceptibility to proliferative
CC disorders, including cancer using a tumor host range mutant (T-HR
CC mutant) virus. The invention also provides the use of Sal2 genes and
CC proteins in methods of identifying a mammal having, or at a risk of
CC acquiring a proliferative disease. T-HR mutants are used to kill cancer
CC cells such as one carrying a Sal2 alteration. Transgenic and knockout
CC mouse comprising Sal2 nucleic acid are useful as research tools to
CC determine genetic and physiological features of cancer and for
CC identifying compounds that can affect ovarian and other tumors. The
CC present sequence is human Sal2 protein mutant (G744R).
CC Note: This sequence is not shown in the specification but is derived
CC from the human Sal2 wild-type protein shown as SEQ ID NO:1 (AAE17954)
CC in page 77-79 of the specification.
XX
SQ Sequence 1005 AA;
Query Match 99.8%; Score 5269; DB 23; Length 1005;
Best Local Similarity 99.9%; Pred. No. 1.2e-294;
Matches 1004; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 MAHSESRSLRGVPAEPAELGDASEEDHPQYCAKCAQFTDPTERLAHONACSTDPPV 60
Db 1 MAHSESRSLRGVPAEPAELGDASEEDHPQYCAKCAQFTDPTERLAHONACSTDPPV 60
Qy 61 MVIIGGQENPNNSASSEPRPGHNNPOYMDTEHSPPOGSSVPTDPTWGERGEES 120
Db 61 MVIIGGQENPNNSASSEPRPGHNNPOYMDTEHSPPOGSSVPTDPTWGERGEES 120
Qy 121 GHFLVAATGTAAAGGGGILLASPKLGATPLPESTPAPPPPPPPPGVSGHLNIPLI 180
Db 121 GHFLVAATGTAAAGGGGILLASPKLGATPLPESTPAPPPPPPPPGVSGHLNIPLI 180
Qy 181 LEELRVLQQRIOHOMQMEQICROYLLLSGQTVAPASPSELPGTASTPDLPLF 240
Db 181 LEELRVLQQRIOHOMQMEQICROYLLLSGQTVAPASPSELPGTASTPDLPLF 240
Qy 241 SPIKPVQTSKTLASSSSSSSSGAETPKQAFHLNPLGSOHPFSGAGGRSHKPTAP 300
Db 241 SPIKPVQTSKTLASSSSSSSSGAETPKQAFHLNPLGSOHPFSGAGGRSHKPTAP 300
Qy 301 SPALPGSTDQLIASPHLAFPSTTGLAAQCIGAARGLAATASGLLKPKNSGELSYGEV 360
Db 301 SPALPGSTDQLIASPHLAFPSTTGLAAQCIGAARGLAATASGLLKPKNSGELSYGEV 360
Qy 361 MGPLEKPGRRKRCFKACVFGSDSALOHLRSHTGRRPKCNVCGRNFTTGRNLKVHFR 420
Db 361 MGPLEKPGRRKRCFKACVFGSDSALOHLRSHTGRRPKCNVCGRNFTTGRNLKVHFR 420
Qy 421 HREKYPPVOMNPHVPEHLDVITSSGLPYGMSVPEPEKAEEEAATPGGVERKPLVASTT 480
Db 421 HREKYPPVOMNPHVPEHLDVITSSGLPYGMSVPEPEKAEEEAATPGGVERKPLVASTT 480
Qy 481 ALSATESLTLSTASGATATAGLPANFKFVLMKAVEPKKADENTPPGSEGAISGVAES 540
Db 481 ALSATESLTLSTASGATATAGLPANFKFVLMKAVEPKKADENTPPGSEGAISGVAES 540
Qy 541 STATLMQSLKMTSLPSMALLTNHFKSTGSPPLPCARALGASPSSTSKLOLVEKIDRQ 600
Db 541 STATLMQSLKMTSLPSMALLTNHFKSTGSPPLPCARALGASPSSTSKLOLVEKIDRQ 600
Qy 601 GAVAVTSAASGAPTTSAAPASSSAGSPNCVICTLRVLSCPRALRLHYGQHGGERPFCK 660
Db 601 GAVAVTSAASGAPTTSAAPASSSAGSPNCVICTLRVLSCPRALRLHYGQHGGERPFCK 660
Qy 661 VCGRAFSTRGNLRAHFVGHKASPAARAONSCPICOKKFTNAVTLQOHVRLHMLGGQIPNGG 720
Db 661 VCGRAFSTRGNLRAHFVGHKASPAARAONSCPICOKKFTNAVTLQOHVRLHMLGGQIPNGG 720
Qy 721 TALEGGGAOAENGSEOSTVSGAGSPPOOQSOQSPPEELSEEEDEDEEDVDDED 780
Db 721 TALEGGGAOAENGSEOSTVSGAGSPPOOQSOQSPPEELSEEEDEDEEDVDDED 780
Qy 781 LAGRSGSGGKALISVRGDSEBASGAEEVGTVAATAATAGKEMDNKKTQOSSLPPPP 840
Db 781 LAGRSGSGGKALISVRGDSEBASGAEEVGTVAATAATAGKEMDNKKTQOSSLPPPP 840

DB 781 LAAGSSGGKAIKSVKSDSEASAEVEGTVAAPATGKENDSNEKTTQOSSLPPPPP 840
QY 841 PDSLDQPPMEQSSGVGGRKEEGKPERSSPASALTEGEATSVTLVELSLQEAMRK 900
DB 841 PDSLDQPPMEQSSGVGGRKEEGKPERSSPASALTEGEATSVTLVELSLQEAMRK 900
QY 901 EPESSSRKACEVCGQAPPSQALAEHOKTHPEGLFTCVFCROGFLERATLKKHMLA 960
DB 901 EPESSSRKACEVCGQAPPSQALAEHOKTHPEGLFTCVFCROGFLERATLKKHMLA 960
QY 961 HHQVQPPAPHPQNIATLIVPGCSPTSTGTLSPPPRKDDPTIP 1005
DB 961 HHQVQPPAPHPQNIATLIVPGCSPTSTGTLSPPPRKDDPTIP 1005
RESULT 4
ID AAM78838 standard; Protein: 1007 AA.
AC AAM78838;
DT 06-NOV-2001 (first entry)
DE Human protein SEQ ID NO 1500.
XX Human: cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorder; arthritis; inflammation.
XX Homo sapiens.
XX WC200157190-A2.
PD 09-AUG-2001.
XX 05-FEB-2001; 2001WO-US04098.
XX 03-FEB-2000; 2000US-0496914.
PR 27-APR-2000; 2000US-0560875.
PR 20-JUN-2000; 2000US-0598075.
PR 19-JUL-2000; 2000US-0620325.
PR 01-SEP-2000; 2000US-0654936.
PR 15-SEP-2000; 2000US-0663561.
PR 20-OCT-2000; 2000US-0693325.
PR 30-NOV-2000; 2000US-0728422.
XX (HYSE-) HXSEQ INC.
PA
XX Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;
PI Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;
XX WPI: 2001-476283/51.
DR N-PSDB: AAK51971.
XX
PT Nucleic acids encoding polypeptides with cytokine-like activities,
XX useful in diagnosis and gene therapy -
PS Claim 20; Page 3790-3792; 6221pp; English.
XX
XX The invention relates to polynucleotides (AAK51456-AAK53435) and the
CC encoded polypeptides (AAM78323-AAK80302) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activity/inhibin growth and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.

CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
CC (AAM8020) are omitted as the relevant pages from the sequence listing
CC were missing at the time of publication.
XX
SQ Sequence 1007 AA;
Query Match 98.0%; Score 5171; DB 22; Length 1007;
Best Local Similarity 99.5%; Pred. No. 5.3e-289;
Matches 985; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 16 GEPALGDASEEDHPQVCAKCAQFTDPTPEFLAHQACSTDPVMVITIGQENPNNSA 75
DB 18 GPSASENDASEEDHPQVCAKCAQFTDPTPEFLAHQACSTDPVMVITIGQENPNNSA 77
QY 76 SSEPRPEGHNNPQVMDTEHSNPPDSSSVPTDTTWGEBRGEESGHTLVATGTAAGG 135
DB 78 SSEPRPEGHNNPQVMDTEHSNPPDSSSVPTDTTWGEBRGEESGHTLVATGTAAGG 137
QY 136 GGLILASPKLIGATPLPEESTPAPPPPPPPPPPGVSGHLNIPILIEBLRVLQORQHOM 195
DB 138 GGLILASPKLIGATPLPEESTPAPPPPPPPPPPPPGVSGHLNIPILIEBLRVLQORQHOM 197
QY 196 QMTQICRQVLLIGLQTVGAPASPELPGTGAASSTKPLLPFSPIKPVQTSKLASS 255
DB 198 QMTQICRQVLLIGLQTVGAPASPELPGTGAASSTKPLLPFSPIKPVQTSKLASS 257
QY 256 SSSSSSSSAETPKQAFHLYHPLGSHPPSAGVGSNKTPTAPSPALPGSTDOLIASP 315
DB 258 SSSSSSSSAETPKQAFHLYHPLGSHPPSAGVGSNKTPTAPSPALPGSTDOLIASP 317
QY 316 HLAFFTGTLLAOCIGAARGLATASPGLLKPNKSGELSYGEVMGPLEKPGRRKCRF 375
DB 318 HLAFFTGTLLAOCIGAARGLATASPGLLKPNKSGELSYGEVMGPLEKPGRRKCRF 377
QY 376 CAKVEGSDALQILHRSHTGERPYKCNVCGNRTTGNLKVHFRHREKTPHYVQNNPHV 435
DB 378 CAKVEGSDALQILHRSHTGERPYKCNVCGNRTTGNLKVHFRHREKTPHYVQNNPHV 437
QY 436 PEHLDYVITSSGLPYGMSVPRKEEAAATPGGVERPLVASTALSTESTLTLSTSA 495
DB 438 PEHLDYVITSSGLPYGMSVPRKEEAAATPGGVERPLVASTALSTESTLTLSTSA 497
QY 496 GTATAPGLPAFNKFVLMKAEPKAKADENTPPGSEGSAISGVAESSTATLMOLSKLMTSL 555
DB 498 GTATAPGLPAFNKFVLMKAEPKAKADENTPPGSEGSAISGVAESSTATLMOLSKLMTSL 557
QY 556 PSMALLTNNHFKSTGSEPLPLCARALGASPSSETSKLQDLVEKIDROGAVAVTSAAGAPTT 615
DB 558 PSMALLTNNHFKSTGSEPLPLCARALGASPSSETSKLQDLVEKIDROGAVAVTSAAGAPTT 617
QY 616 SAPAPSSASSGPNCYICLRVLSCPRALRLHYOGHGERPKCVCGAFTSRGNLRAH 675
DB 618 SAPAPSSASSGPNCYICLRVLSCPRALRLHYOGHGERPKCVCGAFTSRGNLRAH 677
QY 676 FVGHKASPARAQNNSCPICQKFTNAVTLQOHVNRHNLGQILPNGGTALPEGGGAQENG 735
DB 678 FVGHKASPARAQNNSCPICQKFTNAVTLQOHVNRHNLGQILPNGGTALPEGGGAQENG 737
QY 736 EOSTVSGAGSFPOQSOQSPDEBELSEEBEDEDVEDYDDEDSLAGGSEGEKAIS 795
DB 738 EOSTVSGAGSFPOQSOQSPDEBELSEEBEDEDVEDYDDEDSLAGGSEGEKAIS 797
QY 796 VQGDSEASGAEEVGTVAAPATGKENDSNEKTTQOSSLPPPPPPPSLDQPPMEQSS 855
DB 798 VQGDSEASGAEEVGTVAAPATGKENDSNEKTTQOSSLPPPPPPPSLDQPPMEQSS 857
QY 856 GVLGKEEGKPERSSPASALTEGEATSVTLVELSLQEAMRKPPGSSSRKACEVCG 915
DB 858 GVLGKEEGKPERSSPASALTEGEATSVTLVELSLQEAMRKPPGSSSRKACEVCG 917
QY 916 QAFPSQALAEHOKTHPEGLFTCVFCROGFLERATLKKHMLAHQVQPPAPHPQNI 975
DB 918 QAFPSQALAEHOKTHPEGLFTCVFCROGFLERATLKKHMLAHQVQPPAPHPQNI 977

QY 976 AALSLVPGCSPTSITSTGLSPFPRKDDPTIP 1005
| | | | | | | | | | | | | | | | | | | |
DB 978 AALSLVPGCSPTSITSTGLSPFPRKDDPTIP 1007

RESULT 5
AAM79822
ID AAM79822 standard; Protein; 1019 AA.
XX AAM79822;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human protein SEQ ID NO 3468.
XX
KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; hematopoiesis;
KM tissue growth factor; immunomodulatory; cancer; leukaemia;
XX nervous system disorder; arthritis; inflammation.
OS Homo sapiens.
XX
PS WO200157190-A2.
PN
XX
PD 09-AUG-2001.
PE
PF 05-FEB-2001; 2001WO-US04098.
PX
XX
PR 03-FEB-2000; 2000US-0496914.
PR 27-APR-2000; 2000US-0560875.
PR 20-JUN-2000; 2000US-0598075.
PR 19-JUL-2000; 2000US-0620325.
PR 01-SEP-2000; 2000US-0654936.
PR 15-SEP-2000; 2000US-0663561.
PR 20-OCT-2000; 2000US-0693325.
PR 30-NOV-2000; 2000US-0728422.

(HYSE-) HYSEQ INC.
PA
XX
PI Tang YH, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;
PI Zhao QH, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
PI Xue AJ, Yang Y, Wejrtman T, Goodrich R;
XX
DR WPI: 2001-476283/51.
DR N-PSDB: AAK52955.
XX
PT Nucleic acids encoding polypeptides with cytokine-like activities,
PT useful in diagnosis and gene therapy -
XX

Claim 20; Page 347-348; 6221pp; English.

The invention relates to polynucleotides (AAK51456-AAK53435) and the encoded polypeptides (AAM78323-AAK80302) that exhibit activity elating to cytokine, cell proliferation or cell differentiation or which may induce production of other cytokines in other cell populations. The polynucleotides and polypeptides are useful in gene therapy, vaccines or peptide therapy. The polypeptides have various cytokine-like activities, e.g. stem cell growth factor activity, haematopoiesis regulating activity, tissue growth factor activity, immunomodulatory activity and activin/inhibin activity and may be useful in the diagnosis and/or treatment of cancer, leukaemia, nervous system disorders, arthritis and inflammation.

Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666 (AAM80020) are omitted as the relevant pages from the sequence listing were missing at the time of publication.

Sequence 1019 AA;

Query Match 97.8%; Score 5161; DB 22; Length 1019;
Best Local Similarity 99.3%; Pred. No. 2e-208;
Matches 983; Conservative 1; Indels 0; Gaps 0;

QY	16	GEPELIGDADSEEDHDPQVCAKCAQAFPTDEFLAHQNACTDPVAVYIIGGQENPNNSA	75
Db	30	GPSASENDADSEEDHDPQVCAKCAQAFPTDEFLAHQNACTDPVAVYIIGGQENPNNSA	89
QY	76	SSPEPECHNNPQVMDTHSNPDSGSSVPTDPMGPERGESSGHFLVAATGTAAGG	135
Db	90	SSPEPECHNNPQVMDTHSNPDSGSSVPTDPMGPERGESSGHFLVAATGTAAGG	149
QY	136	GGILIASPKLATPLPEPESTPAPPPPPPPPPGGSHLNLPLILEELRYLQOQRIHQ	195
Db	150	GGILIASPKLATPLPEPESTPAPPPPPPPPPGGSHLNLPLILEELRYLQOQRIHQ	209
QY	196	QMTBQICQVLLLSLQGTVAAPASSELPGTGTASTKPLPLPSPIKPVQTSKLASS	255
Db	210	QMTBQICQVLLLSLQGTVAAPASSELPGTGTASTKPLPLPSPIKPVQTSKLASS	269
QY	256	SSSSSSSSGATPPQAFPHLKHPLGSONPFSAGYGRSHKPPPADSPALPGSDQLIASF	315
Db	270	SSSSSSSSGATPPQAFPHLKHPLGSONPFSAGYGRSHKPPPADSPALPGSDQLIASF	329
QY	316	HLAPSTTGLLAAQCLGARGLEATASGGLKPKNGSGELSTGEVWGPLEKGGRRHKCP	375
Db	330	HLAPSTTGLLAAQCLGARGLEATASGGLKPKNGSGELSTGEVWGPLEKGGRRHKCP	389
QY	376	CAKVEGSDSALOHLRSHTGERPYKCNVGNRFETTRGNLKVHFHRRREKYPHYQNNPBY	435
Db	390	CAKVEGSDSALOHLRSHTGERPYKCNVGNRFETTRGNLKVHFHRRREKYPHYQNNPBY	449
QY	436	PEHNDYITSSGLLYGMSVPEPKKEEENATGGGGERKPIVASTALASTESLTLSTSA	495
Db	450	PEHNDYITSSGLLYGMSVPEPKKEEENATGGGGERKPIVASTALASTESLTLSTSA	509
QY	496	GTAATAPGLPAFNKFLVLMKAYEKKKADENTPPGSEGSALISGVAESTATBMQLSKLVTSL	555
Db	510	GTAATAPGLPAFNKFLVLMKAYEKKKADENTPPGSEGSALISGVAESTATBMQLSKLVTSL	569
QY	556	PSMALLTNHKKSTGSPFLPLICARALGASPSSETSKILOLVEKIDROGANAATSAAGAPTT	615
Db	570	PSMALLTNHKKSTGSPFLPLICARALGASPSSETSKILOLVEKIDROGANAATSAAGAPTT	629
QY	616	SAPAPSSSASSGPNOCYICLVLSCLPRALRYLHNGOHGGERPKCVCGRAFSTRGNLRAH	675
Db	630	SAPAPSSSASSGPNOCYICLVLSCLPRALRYLHNGOHGGERPKCVCGRAFSTRGNLRAH	689
QY	676	PVGHKASPARAONSCPTICQKKTNAVTALQOHVNMHLGGQIPNGSTALPEGGAQENGS	735
Db	690	PVGHKASPARAONSCPTICQKKTNAVTALQOHVNMHLGGQIPNGSTALPEGGAQENGS	749
QY	736	EGSTVSGAGSFPOOQSOQRPSEELISEEEDDEDEEDDYDDEDSLARGSSGGEKALS	795
Db	750	EGSTVSGAGSFPOOQSOQRPSEELISEEEDDEDEEDDYDDEDSLARGSSGGEKALS	809
QY	796	VRGDSEBASGAEEBVGTVAAATAGKEMDSNEKTTQOSSLRPPPPDDSLDQPOMEQSSS	855
Db	810	VRGDSEBASGAEEBVGTVAAATAGKEMDSNEKTTQOSSLRPPPPDDSLDQPOMEQSSS	869
QY	856	GVYLGKKEGCKPERSSSPASALTPREGATSVTLVEELSLDAMRKEPESSSRKACEVCG	915
Db	870	GVYLGKKEGCKPERSSSPASALTPREGATSVTLVEELSLDAMRKEPESSSRKACEVCG	929
QY	916	QAFPSOALAEHOKTDPKESGPLFCVYCGROGFLERATLKKMMLAHNOVOFAPRGPONI	975
Db	930	QAFPSOALAEHOKTDPKESGPLFCVYCGROGFLERATLKKMMLAHNOVOFAPRGPONI	989
QY	976	AALSLVPGCSPSITSTGLSPFRKDDPTIP	1005
Db	990	AALSLVPGCSPSITSTGLSPFRKDDPTIP	1019
RESULT 6			
ID	AAEL1955		
XX	AAEL1955 standard; Protein; 1002 AA.		

AC	AAE17955;	
XX		
DT	07-MAY-2002 (first entry)	
XX		
DE	Mouse Sal2 protein.	
XX		
KW	Mouse; proliferative disorder; tumor host range mutant virus; cancer;	
KW	T-HR mutant; Sal2 protein; ovarian tumour.	
XX		
OS	Mus musculus.	
XX		
FH	Key	Location/Qualifiers
FT	Misc-difference 252	
FT	/note= "Encoded by CTC"	
XX		
PN	W0200204596-A2.	
XX		
PD	17-JAN-2002.	
XX		
PF	05-JUL-2001; 2001WO-US21354.	
XX		
PR	07-JUL-2000; 2000US-216723P.	
PR	19-MAR-2001; 2001US-0812471.	
PR	19-MAR-2001; 2001US-0812633.	
XX		
PA	(HARD) HARVARD COLLEGE.	
PA	(BGHM) BRIGHAM & WOMENS HOSPITAL INC.	
XX		
PI	Benjamin TL, Li D, Mok SC, Cramer DW, Ma Y;	
XX		
DR	WPI: 2002-164637/21.	
DR	N-PSDB: AAD28652.	
XX		
PT	Detecting protein involved in susceptibility to proliferative disease,	
PT	by infecting normal and abnormal proliferating cells with mutant virus,	
PT	detecting mutated protein allowing growth of mutant on abnormal cells	
XX		
PS	Disclosure: Page 85-87; 92pp; English.	
XX		
CC	The invention relates to a method for the identification of genes	
CC	and their encoded proteins involved in susceptibility to proliferative	
CC	disorders, including cancer using a tumor host range mutant (T-HR	
CC	mutant) virus. The invention also provides the use of Sal2 genes and	
CC	proteins in methods of identifying a mammal having, or at a risk of	
CC	acquiring a proliferative disease. T-HR mutants are used to kill cancer	
CC	cells such as one carrying a Sal2 alteration. Transgenic and knockout	
CC	mouse comprising Sal2 nucleic acid are useful as research tools to	
CC	determine genetic and physiological features of cancer and for	
CC	identifying compounds that can affect ovarian and other tumours. The	
CC	present sequence is mouse Sal2 protein.	
XX		
SQ	Sequence 1002 AA:	
	Query Match 88.6%; Score 4677.5; DB 23; Length 1002;	
	Best Local Similarity 89.0%; Pred. No. 1.2e-260;	
	Matches 894; Conservative 25; Mismatches 83; Indels 3; Gaps 3;	
OY	1 MAHESRRSLRGVPAEPALGDASEDHQVCAKCAQAFDPTFLAHQNACTDPPV 60	
DB	1 MAQETGSSSRLLGPGCEPARGGDAEENHPQVCAKCAQAFDPTFLAHQNSCCDPPV 60	
OY	61 AVIIGQENPNNSASSEPEEGHNNPOVMDTEHNSNPDSGSSVPPDPTWGPGRGESS 120	
DB	61 MYIIGQENPNNSASASAPPEEGHSRQVMDTEHNSNPDSGSSGPPDPTWGPGRGESS 120	
OY	121 GFLVAATGTAAAGGGGLLASPKLGATPLPEESTPAPPPPPPPPGVSGHLNPLI 180	
DB	121 GGLVATGTAAGGGGLLASPKLGATPLPEESTPAPPPPPPPPGVSGHLNPLI 180	
OY	181 LDELRLQORQHOMQMTEDICRGVLLLSLGQYTGARASPELDETGTASSRKPLPLF 240	
DB	181 LDELRLQORQHOMQMTEDICRGVLLLSLGQYTGARASPELDETGTASSRKPLPLF 240	

OY	241 SPIKPVOTSKTLASSSSSSSGAETPKQAFHLXHPLSQHPSAGVGSHKPTPAP 300	
DB	241 SPIKAQCTKTTA-SSSSSSSSGAETPKQAFHLXHPLSQHPSAGVGSHKPTPAP 299	
OY	301 SPALPGSTDQILASPHLAFPTTGLLAOCLGAARGLBATSPGLLKRKNGSGELSYGY 360	
DB	300 SPALPGSTQILASPHLAFPTTGLLAOCLGAARGLBMAASPGLLKRKNGSGELGYEV 359	
OY	361 MGPLEKPGGRHKCRCAKAVGSDSALOIHLSHNGEPRYKCNVCGNRRFTTRNLKVHFR 420	
DB	360 ISSLEKPGGRHKCRCAKAVGSDSALOIHLSHNGEPRYKCNVCGNRRFTTRNLKVHFR 419	
OY	421 HREKYPHVOMNPNRYPVEHLDVYITSSGLPYGMSVPEPEKAEBAATPGGVERKPLVASTT 480	
DB	420 HREKYPHVOMNPNRYPVEHLDVYITSSGLPYGMSVPEPEKAEBAATPGGVERKPLVASTT 479	
OY	481 ALSATESLTLLSTAGTATAPGLPAEFNKVYLMKAVERKNADENTPPGSEGAISGVAES 540	
DB	480 ALSATESLTLLSTAGTAVAPGLPTFNKFVLMKAVERKNADENTPPGSEGAISGVAES 539	
OY	541 STATLMQLSKLMSTLSPWALLTNHFKSTGSRPLPLCARALGASPSGTSKLOOLVEKIDRQ 600	
DB	540 GSATRMQLSKLVSLSPWALLTNHFKSTGSRPFVYLEPLGASPSGTSKLOOLVEKIDRQ 599	
OY	601 GAVAVTSAASGAPPTAPAPSSSSASGPNOCVICLRVLSCPRALRLHYGOHGERPFKCK 660	
DB	600 GAVAVASTASGAPTTAPAPSSSA-SGPNOCVICLRVLSCPRALRLHYGOHGERPFKCK 658	
OY	661 VCGRAESTRGNLRAHFVGHKASPARAQNSCPICOKKFTNAVTLOOHVBMHLGGQIPNGG 720	
DB	659 VCGRAESTRGNLRAHFVGHKTSPARAQNSCPICOKKFTNAVTLOOHVBMHLGGQIPNGG 718	
OY	721 TALPEGGGAAGENGSRQSTVSGAGSFPOQOOSQPSPEELSEEEDEDEEDVDYDEDS 780	
DB	719 SALSSEGGGAAGENGSRQSTVSGAGSFPOQOQPSPEEEM-EEEDDEDEEDVDYDEDS 777	
OY	781 LAGRGSESGEKAISVNGDSEASGAEEVGTVAATAGKEMDSNEKTTQOSSLPPPPP 840	
DB	778 LAGRGSESGEKAISVNGDSEEVSGAEEVATSVAAPTTKEMDSNEKAPQHTLPPPPP 837	
OY	841 PSLDQPPQMEGSSGYLGKKEGKPEKRRSSPASALTPREGATSVTLVEELSLQEMAKK 900	
DB	838 PNLDHPQPEMEGCTSDVSGAMEEBAKLEGISSPMALTOEGGCTSTPLVEELNPEAMKK 897	
OY	901 EPGESSRKAACEVCGAGFPQOALAEHOKTHPKREGPLFTVCFCROGFLERATLKKHMLA 960	
DB	898 DPESSGRKACEVCGGGSFPTOTALAEHOKTHPKDGPLFTVCFCROGFLDRATLKKHMLA 957	
OY	961 HHQVPPFAPHPQONIALSLVPGCSPSITSTGLSPFPRKDDPTIP 1005	
DB	958 HHQVPPFAPHPQONIALSLVPGCSSSIPSPGLSPFPRKDDPTMP 1002	
RESULT 7		
AAB93193		
ID	AAB93193 standard; Protein: 813 AA.	
XX		
AC	AAB93193;	
DT	26-JUN-2001 (first entry)	
XX		
DE	Human protein sequence SEQ ID NO:12150.	
KW	Human; primer; detection; diagnosis; antisense therapy; gene therapy.	
OS	Homo sapiens.	
XX		
PN	EP1074617-A2.	
XX		
PD	07-FEB-2001.	
XX		
PF	28-JUL-2000; 2000EP-0116126.	

QY 839 P-----PPDSL----- 844
Db 832 PNLILMAAREEMALGNHNAKRLPLPRLPGFMGLHPPRVNCLCFKMLPSLALESIL 891
QY 845 -----DQPMEOGSSGVLGK-----EEGK-PERSSS 872
Db 892 QSEHAKEPATGHAQRQCSDAQSP--YGAKLTILPNLFAKKRPSSSSSGEKLPRESSNP 949
QY 873 PASALTP-----EGEATSVTL----- 888
Db 950 PFPANENPATPTIKEDPDDEQLVVEGASAGSGGTATSNVPEAGDAEOSLMMQILNAN 1009
QY 889 ---VEELSLQEA-MRKEPESS-----SRKACEVCGOAFPPSOALEHOKTHPKESPLF 938
Db 1010 RFPASPLDFOQLMAGSPRTSSLDPPVNNKHFCVCHCRNFSSSSALQIHMKRTHTGDKP-F 1068
QY 939 TCVCRCQGLERATLKKHMLLAHQVQFPAPRGPNITATSLVPG 983
Db 1069 QCNVCKAKFTTKGNLKVHM-GTHMWTNPTSRGRARMSELPLMRPG 1112
RESULT 11
AAB92796
ID AAB92796 standard; Protein; 200 AA.
AC AAB92796:
DT 26-JUN-2001 (first entry)
DE Human protein sequence SEQ ID NO:11298.
XX Human; primer: detection; diagnosis; antisense therapy; gene therapy.
OS Homo sapiens.
XX EP1074617-A2.
PN 07-FEB-2001.
PD 28-JUL-2000; 2000EP-0116126.
PF 29-JUL-1999; 99JP-0248036.
PR 27-AUG-1999; 99JP-0300253.
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX (HELI-) HELIX RES INST.
PA Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
DR WPI: 2001-318749/34.
XX WPI: 2001-318749/34.
PT Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX Claim 8; SEQ ID 11298; 2537pp + CD ROM; English.
XX The present invention describes primer sets for synthesizing 5602
CC full-length cDNAs defined in the specification, where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of

CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX
SQ Sequence 200 AA:
Query Match 11.3%; Score 596; DB 22; Length 200;
Best Local Similarity 92.4%; Pred. No. 7.9e-27;
Matches 109; Conservative 0; Mismatches 9; Indels 0; Gaps 0;
QY 16 GEPALGCDASEEDHPQVCAKCAQFTDPTFEFLAHQNACTDPYVVIIGQENPNSSA 75
Db 18 GPSASENGDASEEDHPQVCAKCAQFTDPTFEFLAHQNACTDPYVVIIGQENPNSSA 77
QY 76 SSEPRPEGHNPNQVMDTEHNSNPPDSGSSVPTDPTWGPERRGEESGHFLVAATGTAAG 133
Db 78 SSEPRPEGHNPNQVMDTEHNSNPPDSGSSVPTDPTWGPERRGEESGHFLVAATEPYCG 135
RESULT 12
AAB30894
ID AAB30894 standard; Peptide; 330 AA.
XX AAB30894:
AC AAB30894:
DT 01-FEB-2002 (first entry)
DE Peptide #3545 encoded by breast cell single exon nucleic acid probe.
XX Human; microarray: single exon probe; gene expression; breast;
KW disease; cancer.
XX Homo sapiens.
OS WO200157271-A2.
XX 09-AUG-2001.
PD 30-JAN-2001; 2001WO-US00662.
PF 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
PA Penn SG, Hanzel DK, Chen W, Rank DR;
PI WPI: 2001-496933/54.
XX WPI: 2001-496933/54.
XX New spatially-addressable set of single exon nucleic acid probes,
PT useful for measuring gene expression in sample derived from human
PT breast, comprises number of single exon nucleic acid probes -
XX Claim 27; SEQ ID NO 13862; 327pp + sequence listing; English.
XX The invention relates to a spatially-addressable set of single exon
CC nucleic acid probes for measuring gene expression in a sample derived
CC from human breast and BT 474 cells. The method involves contacting
CC the probes with a collection of detectably labelled nucleic acids
CC derived from mRNA of human breast, and then measuring the label

KM epilepsy; cancer.
XX
OS Homo sapiens.
XX
PN WO200157275-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00667.
XX
PR 04-FEB-2000; 2000US-0180312.
XX
PR 26-MAY-2000; 2000US-0207456.
XX
PR 30-JUN-2000; 2000US-0608408.
XX
PR 03-AUG-2000; 2000US-0632366.
XX
PR 21-SEP-2000; 2000US-0234687.
XX
PR 27-SEP-2000; 2000US-0236359.
XX
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-483446/52.
XX
PT Single exon nucleic acid probes for analyzing gene expression in human
XX brains -
XX
PS Example 4; SEQ ID NO: 28963; 650pp + Sequence Listing; English.
XX
CC The present invention provides a number of single exon nucleic acid
XX probes which are derived from genomic sequences expressed in the human
XX brain. They can be used to measure gene expression in brain cell samples,
XX which may enable the diagnosis and improved treatment of nervous system
XX diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
XX epilepsy and cancers. The present sequence is a protein encoded by one of
XX the probes of the invention.
XX
SQ Sequence 330 AA;
XX
Query Match 9.3%; Score 489; DB 22; Length 330;
Best Local Similarity 33.3%; Pred. No. 2,1e-20;
Matches 186; Conservative 51; Mismatches 114; Indels 108; Gaps 14;
XX
OY 529 SEGSAISGVAESSTATIMQLSKMLTSLPSWALITNHFKSTGSPFLPLCARALGASPSETS 588
DB 3 SSPAADCGPAGSATTTFTNPL-----LP---LMSEQFKA---KPPFGGLDLSAQASSETS 49
XX
OY 589 KLOQLVEKIDROGAVAVTSAASGAPTTAPAPSSASSGPNQVCYLRLVSCPRALRLHY 648
DB 50 KLOQLVENDDK-----ATDPNCCITCHRVLSQCSALAKMHY 85
XX
OY 649 GGGGRRPCKVCVCGRAFTRGNLRAHFVGHKASPAARAQNSCPICQKFTNAVTIQQOHV 708
DB 86 RHTGTGERPFCKICIGRAFTTKGNLKTHTYSVRAMPPLRVQHSQPCICQKFTNAVTIQQOHV 145
XX
OY 709 RHTGTGTGTPNGGTALPREGGAQENGSEOSTVSGAGSPPOQSOOPSPBEELSEEBEED 768
DB 146 RHMHGQIIPN--TPVPDYSSESM-----SDTGSF-----DEKNFDD 180
XX
OY 769 EEBEEDVTDEDSLARGSESGEKAISVRGDESEAGAEVEGTVAATAATACKEMPSNEK 828
DB 181 LDNFSDENMED--CPBGSIPDTPKSAQSDSLSSPLPLEMSSTIALENNQKMINAGLA 238
XX
OY 829 TTQGSLLPPPPDSDLQPOPMEOGS-----SGVLG---KEGGKPERSSSPAS- 875
DB 839 EQLQASL-----KSVENGSIEGDVLITNDSSVSGGMESQASQSAISESTSSM 286
XX
OY 876 -ALTEGGEATSVTLVEELSLQELAMRKPEGSSSRKACEVCGAPPSQAA 923
DB 287 QALSPSNS-----TQEFHKSPTIEKP-----QRAVPSSEA 317
XX
RESULT 15

AAM69245
ID AAM69245 standard; Protein; 330 AA.
XX
AC AAM69245;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human bone marrow expressed probe encoded protein SEQ ID NO: 29551.
XX
KW Human; bone marrow expressed exon; gene expression analysis; probe;
XX microarray; cancer; leukaemia; lymphoma; myeloma.
XX
OS Homo sapiens.
XX
PN WO200157276-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00668.
XX
PR 04-FEB-2000; 2000US-0180312.
XX
PR 26-MAY-2000; 2000US-0207456.
XX
PR 30-JUN-2000; 2000US-0608408.
XX
PR 03-AUG-2000; 2000US-0632366.
XX
PR 21-SEP-2000; 2000US-0234687.
XX
PR 27-SEP-2000; 2000US-0236359.
XX
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-488900/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human bone marrow -
XX
PS Example 4; SEQ ID NO: 29551; 658pp + Sequence Listing; English.
XX
CC The present invention provides a number of single exon nucleic acid
XX probes which are derived from genomic sequences expressed in the human
XX bone marrow. They can be used to measure gene expression in bone marrow
XX samples, which may enable the improved diagnosis and treatment of cancers
XX such as lymphoma, leukaemia and myeloma. The present sequence is a
XX protein encoded by one of the probes of the invention.
XX
SQ Sequence 330 AA;
XX
Query Match 9.3%; Score 489; DB 22; Length 330;
Best Local Similarity 33.3%; Pred. No. 2,1e-20;
Matches 136; Conservative 51; Mismatches 114; Indels 108; Gaps 14;
XX
OY 529 SEGSAISGVAESSTATIMQLSKMLTSLPSWALITNHFKSTGSPFLPLCARALGASPSETS 588
DB 3 SSPAADCGPAGSATTTFTNPL-----LP---LMSEQFKA---KPPFGGLDLSAQASSETS 49
XX
OY 589 KLOQLVEKIDROGAVAVTSAASGAPTTAPAPSSASSGPNQVCYLRLVSCPRALRLHY 648
DB 50 KLOQLVENDDK-----ATDPNCCITCHRVLSQCSALAKMHY 85
XX
OY 649 GGGGRRPCKVCVCGRAFTRGNLRAHFVGHKASPAARAQNSCPICQKFTNAVTIQQOHV 708
DB 86 RHTGTGERPFCKICIGRAFTTKGNLKTHTYSVRAMPPLRVQHSQPCICQKFTNAVTIQQOHV 145
XX
OY 709 RHTGTGTGTPNGGTALPREGGAQENGSEOSTVSGAGSPPOQSOOPSPBEELSEEBEED 768
DB 146 RHMHGQIIPN--TPVPDYSSESM-----SDTGSF-----DEKNFDD 180
XX
OY 769 EEBEEDVTDEDSLARGSESGEKAISVRGDESEAGAEVEGTVAATAATACKEMPSNEK 828
DB 181 LDNFSDENMED--CPBGSIPDTPKSAQSDSLSSPLPLEMSSTIALENNQKMINAGLA 238
XX
OY 829 TTQGSLLPPPPDSDLQPOPMEOGS-----SGVLG---KEGGKPERSSSPAS- 875

[illegible]

Oy	649	GONGGEPKCKYAGRASTGTGNLRHAFVGHKASPAARONSOPTLOKKFTTNAVTTLOQH	708
Db	86	RTHGGEPFKCTIKGRATTKGNLTHTSVHRAMPRLRHOHSCLPQCKFTTNVVLQHI	145
Oy	709	RMLLGOIPNGGTLPRCGAAGAEENGSEOSTIVSGASFPOOQSOPSPBELSEEBEED	768
Db	146	RMHHGGIIPN--TPVPDYSSEME-----SDTGSE-----DEKNFDD	180
Oy	769	EEREEVDYTDSDLAGRSGSEGEKAISVRGDSFEASGAEEYGVAAAATACKEMDNSNEK	828
Db	181	LDRNFSDNMED--CPESIDPTPKRSADASODSLSSPLPLEMSSIATLNQMIMINAGLA	238
Oy	829	TFOOSSLPPEPPPSLDQPOMEOS-----SGVLGG--KEGGKPERSSSPAS-	875
Db	239	EQLQASL-----KSVENGSITGDVLTNDSSSYVGDMESOSAGSPAISETSTM	286
Oy	876	-ALTPPEGATSVTLVEELSLQEAARKPEGESSRRKACEVCQAFPQA	923
Db	287	QALSPSNS-----TOEFHKSPSIEKP-----QRAVPEFA	317
RESULT 17			
ID	AAM29568	standard; Protein: 330 AA.	
XX	AC	AAM29568;	
XX	DT	17-OCT-2001 (first entry)	
DE	XX	Peptide #3605 encoded by probe for measuring placental gene expression.	
XX	KW	Probe: microarray; human; placenta; antenatal diagnosis;	
OS	XX	genetic disorder.	
PN	XX	Homo sapiens.	
PD	XX	WO200157272-A2.	
PF	XX	09-AUG-2001.	
XX	XX	30-JAN-2001; 2001MO-US00663.	
PR	XX	04-FEB-2000; 2000US-0180312.	
PR	XX	26-MAY-2000; 2000US-0207456.	
PR	XX	30-JUN-2000; 2000US-0608408.	
PR	XX	03-AUG-2000; 2000US-0632366.	
PR	XX	21-SEP-2000; 2000US-0234687.	
PR	XX	27-SEP-2000; 2000US-0236359.	
PR	XX	04-OCT-2000; 2000GB-0024263.	
PA	XX	(MOLE-) MOLECULAR DYNAMICS INC.	
PI	XX	Penn SG, Hanzel DK, Chen W, Rank DR;	
DR	XX	WPI; 2001-488897/53.	
PT	XX	Human genome-derived single exon nucleic acid probes useful for	
XX	XX	analyzing gene expression in human placenta -	
PS	XX	Claim 27; SEQ ID No 29837; 654bp; English.	
CC	XX	The present invention relates to single exon nucleic acid probes (SENPs;	
CC	XX	see AI31315-AI57546). The present sequence is a peptide encoded by one	
CC	XX	such probe. The probes are useful for producing a microarray for	
CC	XX	predicting, measuring and displaying gene expression in samples derived	
CC	XX	from human placenta. The probes are useful for antenatal diagnosis of	
CC	XX	human genetic disorders.	
SQ	Sequence	330 AA.	
Query Match	9.3%;	Score 489; DB 22; Length 330;	
Best Local Similarity	33.3%;	Pred No. 2.1e-20;	


```
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX PA Penn SG, Hanzel DK, Chen W, Rank DR;
XX PI WPI; 2001-483446/52.
XX DR
XX PT Single exon nucleic acid probes for analyzing gene expression in human
XX PT brains -
XX
XX Example 4; SEQ ID NO: 30291; 650pp + Sequence Listing; English.
XX
XX The present invention provides a number of single exon nucleic acid
XX probes which are derived from genomic sequences expressed in the human
XX brain. They can be used to measure gene expression in brain cell samples,
XX which may enable the diagnosis and improved treatment of nervous system
XX diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
XX epilepsy and cancers. The present sequence is a protein encoded by one of
XX the probes of the invention.
XX
XX Sequence 336 AA; |
XX
XX Query Match 8.4%; Score 441; DB 22; Length 336;
XX Best Local Similarity 34.3%; Pred. No. 1.2e-17;
XX Matches 119; Conservative 41; Mismatches 89; Indels 98; Gaps 11;
XX
OY 585 SETSKLOOLEKIDROGAVAVTSAASGAPTTAPAPSSSSAGPNOCVICLRVLCPRAL 644
DB 62 SETLKIQLQVLENDIK-----ATDPNCECLICHRVLSGSSSL 97
OY 645 RLHYGOGHGERPRCKVCYKGRAFSTRGNLRHFVGHKASPAARAQNSCPICQKKFTNAVTL 704
DB 98 KMHYRTHYTGGERPRFQCKICGRAFSYKGNLKTGLGVHRTNTSIKTHSCPICQKKFTNAVML 157
OY 705 QOHVRMHLGQIPIPGCTALPBGGAQAQENGSPQSTVSGAGSPPOQSOQSPPEEELSEEE 764
DB 158 QOHIRRMHMGQIIPN--TPLE--NPCDFTGSEPMTVGENGS-----TGAICHDH 202
OY 765 EEEDEEEEDVTDDEDSLGRGSESGGEKALISVRGDSEASGAEEVGTV-----AAATA 819
DB 203 VIESIDVEE-----VSSQEARPSSSSKVPPLPSTHSASPTL 238
OY 820 GKEMDSNEKTTQOSSLPP-----PPPSLDOPQPMEOG-----SSGVLGKEEGG 865
DB 239 GFAM-----MASLIDAPGKVGPAFPNLIQROGSRENGSVESDGLTNDSSSLMGDQE--- 287
OY 866 KPERSSSPASALTPEGEATSVTLIVEELSLQAMRKEPGESSSRKACE 912
DB 288 --YQSRSPDILETTTFQALSPANSQAESIKS---KSPDAGKAESSE 329
XX
XX RESULT 23
XX ID AAM70641 standard; Protein; 336 AA.
XX AC AAM70641;
XX
XX 06-NOV-2001 (first entry)
XX
XX Human bone marrow expressed probe encoded protein SEQ ID NO: 30947.
XX
XX Human: bone marrow expressed exon; gene expression analysis; probe;
XX microarray; cancer; leukaemia; lymphoma; myeloma.
XX
XX Homo sapiens.
XX
XX WO200157276-A2.
XX
XX
```

```
XX
XX 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US00668.
XX
XX 04-FEB-2000; 2000US-0180312.
XX 26-MAY-2000; 2000US-0207456.
XX 30-JUN-2000; 2000US-0608408.
XX 03-AUG-2000; 2000US-0632366.
XX 21-SEP-2000; 2000US-0234687.
XX 27-SEP-2000; 2000US-0236359.
XX 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX PA Penn SG, Hanzel DK, Chen W, Rank DR;
XX PI WPI; 2001-488900/53.
XX DR
XX
XX Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human bone marrow -
XX
XX Example 4; SEQ ID NO: 30947; 658bp + Sequence Listing; English.
XX
XX The present invention provides a number of single exon nucleic acid
XX probes which are derived from genomic sequences expressed in the human
XX bone marrow. They can be used to measure gene expression in bone marrow
XX samples, which may enable the improved diagnosis and treatment of cancers
XX such as lymphoma, leukaemia and myeloma. The present sequence is a
XX protein encoded by one of the probes of the invention.
XX
XX Sequence 336 AA;
XX
XX Query Match 8.4%; Score 441; DB 22; Length 336;
XX Best Local Similarity 34.3%; Pred. No. 1.2e-17;
XX Matches 119; Conservative 41; Mismatches 89; Indels 98; Gaps 11;
XX
OY 585 SETSKLOOLEKIDROGAVAVTSAASGAPTTAPAPSSSSAGPNOCVICLRVLCPRAL 644
DB 62 SETLKIQLQVLENDIK-----ATDPNCECLICHRVLSGSSSL 97
OY 645 RLHYGOGHGERPRCKVCYKGRAFSTRGNLRHFVGHKASPAARAQNSCPICQKKFTNAVTL 704
DB 98 KMHYRTHYTGGERPRFQCKICGRAFSYKGNLKTGLGVHRTNTSIKTHSCPICQKKFTNAVML 157
OY 705 QOHVRMHLGQIPIPGCTALPBGGAQAQENGSPQSTVSGAGSPPOQSOQSPPEEELSEEE 764
DB 158 QOHIRRMHMGQIIPN--TPLE--NPCDFTGSEPMTVGENGS-----TGAICHDH 202
OY 765 EEEDEEEEDVTDDEDSLGRGSESGGEKALISVRGDSEASGAEEVGTV-----AAATA 819
DB 203 VIESIDVEE-----VSSQEARPSSSSKVPPLPSTHSASPTL 238
OY 820 GKEMDSNEKTTQOSSLPP-----PPPSLDOPQPMEOG-----SSGVLGKEEGG 865
DB 239 GFAM-----MASLIDAPGKVGPAFPNLIQROGSRENGSVESDGLTNDSSSLMGDQE--- 287
OY 866 KPERSSSPASALTPEGEATSVTLIVEELSLQAMRKEPGESSSRKACE 912
DB 288 --YQSRSPDILETTTFQALSPANSQAESIKS---KSPDAGKAESSE 329
XX
XX RESULT 24
XX ID AAM18485 standard; Protein; 336 AA.
XX AC AAM18485;
XX
XX 12-OCT-2001 (first entry)
XX
XX Peptide #4919 encoded by probe for measuring cervical gene expression.
XX Probe; human; microarray; gene expression; cervical epithelial cell;
XX
XX
```


KM cervical cancer.
XX Homo sapiens.
OS WO200157278-A2.
PN
XX
XX 09-AUG-2001.
PD
XX
XX 30-JAN-2001; 2001WO-US00670.
PF
XX
XX 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR:
XX WPI; 2001-488901/53.
XX
XX Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human cervical epithelial cells -
XX
XX Claim 27; SEQ ID No 23311; 487bp; English.
PS
XX
XX The present invention relates to human single exon nucleic acid probes
CC (SENPs; see A110068-A18459). The present sequence is a peptide encoded
CC by one such probe. The SENPs are derived from human HeLa cells. The SENPs
CC can be used to produce a single exon microarray, which can be used for
CC measuring human gene expression in a sample derived from human cervical
CC epithelial cells. By measuring gene expression, the probes are therefore
CC useful in grading and/or staging of diseases of the cervix, notably
CC cervical cancer.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 336 AA:
SQ
Query Match 8.4%; Score 441; DB 22; Length 336;
Best Local Similarity 34.3%; Pred. No. 1.2e-17;
Matches 119; Conservative 41; Mismatches 89; Indels 98; Gaps 11;
OY 585 SETSKLOQLVEKIDROGAVAVTSAASGAPPTTSAPAPSSASSGPNQVCYICLRVLSCPRAL 644
DB 62 SETKLQQLVENIDK-----ATTDPECLICHRLVLSQSSSL 97
OY 645 RLHYGOHGERPPKCKVCGRAFTSGNLRHAFVGHKASPARAONSCPTCKQKFTNAVTL 704
DB 98 KMHYRTHNGERPFQCKICGRAFTSGNLRHAFVGHKASPARAONSCPTCKQKFTNAVTL 157
OY 705 OOHVAMHLAGQIPNGGTALPEGGAAGAEQSGSTVSGAGSPPOQSOQSPPEEELSEEE 764
DB 158 OOHIRMHMGQIPN--TPLE--NPCDFTGSEPMIVGENG-----TGAICHDD 202
OY 765 EEEDEEEDVDYDSDSLAGRSGESGGEKAISVRGDSSEASGAEEVGTV-----AAATA 819
DB 203 VIESIDVEE-----VSSQEAPESSSKVPTPLPSIHSASPTL 238
OY 820 GKEMDSNEKTTQOSSLPPP-----PPPSLDOPQPMEOG-----SSGVLGKKEEG 865
DB 239 GFAM-----MASLDAPGKVGAPAPFNLDQSGRNGSVESDGLTNDSSSLMGDOE--- 287
OY 866 KPERSSPASALTPPEGATSVTLVEELSLQEAAMRKEPGESSSRKACE 912
DB 288 ---YQSHSPDILETTTSFOALSPANSQAESIKS---KSPDAGKAESSE 329
RESULT 25

AAM30956
ID AAM30956 standard; Protein; 336 AA.
XX
XX AAM30956;
AC
XX
XX 17-OCT-2001 (first entry)
DT
XX
XX Peptide #4993 encoded by probe for measuring placental gene expression.
DE
XX
XX Probe; microarray; human; placenta; antenatal diagnosis;
KW genetic disorder.
KW Homo sapiens.
XX
XX WO200157272-A2.
PN
XX
XX 09-AUG-2001.
PD
XX
XX 30-JAN-2001; 2001WO-US00663.
PF
XX
XX 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR:
XX WPI; 2001-48897/53.
XX
XX Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human placenta -
XX
XX Claim 27; SEQ ID No 31225; 654bp; English.
PS
XX
XX The present invention relates to single exon nucleic acid probes (SENPs;
CC see A131315-A157546). The present sequence is a peptide encoded by one
CC such probe. The probes are useful for producing a microarray for
CC predicting, measuring and displaying gene expression in samples derived
CC from human placenta. The probes are useful for antenatal diagnosis of
CC human genetic disorders.
XX
XX Sequence 336 AA:
SQ
Query Match 8.4%; Score 441; DB 22; Length 336;
Best Local Similarity 34.3%; Pred. No. 1.2e-17;
Matches 119; Conservative 41; Mismatches 89; Indels 98; Gaps 11;
OY 585 SETSKLOQLVEKIDROGAVAVTSAASGAPPTTSAPAPSSASSGPNQVCYICLRVLSCPRAL 644
DB 62 SETKLQQLVENIDK-----ATTDPECLICHRLVLSQSSSL 97
OY 645 RLHYGOHGERPPKCKVCGRAFTSGNLRHAFVGHKASPARAONSCPTCKQKFTNAVTL 704
DB 98 KMHYRTHNGERPFQCKICGRAFTSGNLRHAFVGHKASPARAONSCPTCKQKFTNAVTL 157
OY 705 OOHVAMHLAGQIPNGGTALPEGGAAGAEQSGSTVSGAGSPPOQSOQSPPEEELSEEE 764
DB 158 OOHIRMHMGQIPN--TPLE--NPCDFTGSEPMIVGENG-----TGAICHDD 202
OY 765 EEEDEEEDVDYDSDSLAGRSGESGGEKAISVRGDSSEASGAEEVGTV-----AAATA 819
DB 203 VIESIDVEE-----VSSQEAPESSSKVPTPLPSIHSASPTL 238
OY 820 GKEMDSNEKTTQOSSLPPP-----PPPSLDOPQPMEOG-----SSGVLGKKEEG 865
DB 239 GFAM-----MASLDAPGKVGAPAPFNLDQSGRNGSVESDGLTNDSSSLMGDOE--- 287
OY 866 KPERSSPASALTPPEGATSVTLVEELSLQEAAMRKEPGESSSRKACE 912

Db 288 --YQSRSPDILETTSFQALSPANQAESIKS---KSPDAGSKAESSE 329

RESULT 26

AA06075

ID AA06075 standard; Protein; 336 AA.

AC AA06075;

DT 09-OCT-2001 (first entry)

DE Peptide #4757 encoded by probe for measuring breast gene expression.

KW Probe; human; breast disease; breast cancer; development disorder; inflammatory disease; proliferative breast disease; non-carcinoma tumour.

OS Homo sapiens.

PN WO200157270-A2.

PD 09-AUG-2001.

XX 29-JAN-2001; 2001WO-0500661.

XX 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

PI Penn SG, Hanzel DK, Chen W, Rank DR;

DR WPI: 2001-476286/51.

XX Novel single exon nucleic acid probe used to measuring gene expression in a human breast -

PS Claim 27; SEQ ID No 14815; 322pp; English.

XX The present invention relates to novel single exon nucleic acid probes (see AA100010-AA110067). The present sequence is a peptide encoded by one such probe. The probes are useful for measuring human gene expression in a human breast sample, where the probe hybridises at high stringency to a nucleic acid expressed in the human breast. The probes are useful for predicting, diagnosing, grading, staging, monitoring and prognosing diseases of the human breast, particularly those diseases with polygenic aetiology. The diseases include: breast cancer, disorders of development, inflammatory diseases of the breast, fibrocystic changes, proliferative breast disease and non-carcinoma tumours.

CC Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 336 AA:

Query Match 8.4%; Score 441; DB 22; Length 336; Best Local Similarity 34.3%; Pred. No. 1.2e-17; Matches 119; Conservative 41; Mismatches 89; Indels 98; Gaps 11;

QY 585 SETSKLOQVEKIDRGAVAVTSAAGAPRTSAPAPSSASGPMQVCYLRLSCPRAL 644

Db 62 SETKLQQLQVENVIDK-----ATDPNECLTCHRLSCQSSL 97

QY 645 RLHYGQHGGERPFCCKVCGRAFSTRGNLRAHFVGHKASPAARAONSCPTCKKFTNAVTL 704

Db 98 KMYHYHTGGERPQCKICGAFSTKGNLKHHLGVHRTNTSIKTQNSCPICQCKKFTNAVNL 157

QY 705 QQHVMHLLGGQIPNGSTALPEGGGAOENSGESTVSGAGSFPQOOSQOPSPEELSEEE 764

Db 158 QQHIRMHGGQIPN--TLPE--NCDFTGSEPMYTGENS-----TCAICDD 202

QY 765 EEDEEEEDVTDSDSLAGRSGSEGEKAI SVRGDSEEAAGAEVGTV-----AAATA 819

Db 203 VIESIDVEE-----VSSQEAAPSSSKVPTPLPSIHASPTL 238

QY 820 GKENDSNEKTTQOSSLPP-----PPDSLDQPOPMQG-----SSGVLAGKEGG 865

Db 239 GFAM-----MASLDAPGKYGPAPFNLQROGSRENGSVESDGLTNDSSSLMGDOE--- 287

QY 866 KPERSSPASALTPEGFATSVTLVEELSLQEA MRKEGESSRKACE 912

Db 288 --YQSRSPDILETTSFQALSPANQAESIKS---KSPDAGSKAESSE 329

RESULT 27

ABG40328

ID ABG40328 standard; Peptide; 336 AA.

AC ABG40328;

XX 19-AUG-2002 (first entry)

DE Human peptide encoded by genome-derived single exon probe SEQ ID 29993.

XX Human; single exon probe; asthma; lung cancer; COPD, IHD;

KW Chronic obstructive pulmonary disease; interstitial lung disease; familial idiopathic pulmonary fibrosis; neurofibromatosis; tuberous sclerosis; Gaucher's disease; Niemann-Pick disease; Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemosiderosis; pulmonary histiocytosis; lymphangioleiomyomatosis; Kargener syndrome; pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia; primary ciliary dyskinesia; pulmonary hypertension;

KW hyaline membrane disease.

OS Homo sapiens.

XX WO200186003-A2.

PN 15-NOV-2001.

XX 30-JAN-2001; 2001WO-0500665.

XX 04-FEB-2000; 2000US-180312P.

PR 26-MAY-2000; 2000US-207456P.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-234687P.

PR 27-SEP-2000; 2000US-236359P.

PR 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI: 2002-114183/15.

XX Spatially-addressable set of single exon nucleic acid probes, used to measure gene expression in human lung samples -

PS Claim 27; SEQ ID No 29993; 634pp; English.

XX The invention relates to a spatially-addressable set of single exon nucleic acid probes for measuring gene expression in a sample derived from human lung comprising single exon nucleic acid probes having one of 12614 nucleic acid sequences mentioned in the specification, or their complements or the 12387 open reading frames derived from the 12614 probes. Also included are a microarray comprising the novel set of probes; the novel set of probes which hybridise at high stringency to a nucleic acid expressed in the human lung; measuring gene expression in a sample derived from human lung, comprising (a) contacting the array with a collection of detectably labeled nucleic acids derived from human lung

Db	481	CEEQCKGFSWRSLHMHQVRHSGEKPKP	----	CEQCKDSFSAIDFRVHORVHTGKEPK	536
OY	719	GGTALPEGGGAAGNENSGEOSTVSGAGSFPPQOQSOQPSPEELSEEEDEDEEDVTDE	778		
Db	537	CGVC- --GKGFSSQSLQSHQVRHTGKPY	-----	-----	-K 564
OY	779	DSLGRGSESGGEKALSVNRDSEASGAEEVGVYAAATACKEMNSNKTQOSSLPP	838		
Db	565	CDVCKGKGRYSQFIYHQKHGHTGKPKCECG	-----	KGFRSL	----- 604
OY	839	PPPSLDLPQPMEDQSSGVGKKEGKPERSSPSALTPEGEATSVTLVEELSLQEM	898		
Db	605	-----NLRRHQVRHTGKPKPH	-CEQCK	-----	AFSLSNL 634
OY	899	RKEGESSSRK--ACGVCGQAFPSQALAEHQKTHPKGCPPLTCVFCROGFLERAILKKH	956		
Db	635	RVHLGVHTRERKLEPKCECGKGFSSQARLEAHQVRHTGKPY	KCIDCDKDFRRHSRLTYH	693	
RESULT	29				
AAB21003	ID	AAB21003 standard; Protein; 799 AA.			
AC	AAB21003;				
XX	19-DEC-2000	(first entry)			
DT					
XX					
DE					
XX					
KW		Human nucleic acid-binding protein; NuBP; agonist; antagonist; EST; expressed sequence tag; drug screening; recombinant expression; antibody; reproductive disorder; infertility; immunological disorder; neurological disorder; cell proliferative disorder; cancer; tumour.			
XX					
OS		Homo sapiens.			
XX					
PN		WO200044900-A2.			
XX					
PD		03-AUG-2000.			
XX					
PF		28-JAN-2000; 2000WO-US02237.			
XX					
PR		29-JAN-1999; 99US-0117904.			
XX		29-JAN-1999; 99US-0117905.			
XX					
PA		(INCY-) INCYTE PHARM INC.			
XX					
PI		Tang YT, Lal P, Hillman JL, Yue H, Azimzai Y, Lu AMD, Baughn MR;			
PI		Tran B, Shih IL, Au-Young JL;			
DR		WI: 2000-499332/44.			
XX		N-PSDB; AAA72386.			
XX					
PT		Novel nucleic acid binding proteins, used to identify agonists and			
PT		antagonists of them, for the treatment of reproductive, immunological,			
PT		neurological and cell proliferative disorders including cancer			
XX					
PS		Claim 1; Page 98-100; 180pp; English.			
XX					
CC		Sequences AAB20997-B21051 represent novel human nucleic acid-binding			
CC		proteins (NuBAPs) which are encoded by the cDNA sequences			
CC		AAA72382-A72436. The cDNAs were produced by extension from an appropriate			
CC		EST (expressed sequence tag) using primers designed using the EST. The			
CC		invention also relates to expression constructs, host cells and			
CC		transgenic organisms comprising a human NuBP nucleic acid, recombinant			
CC		production of the human NuBAPs, and antibodies against the human NuBAPs,			
CC		and also to methods of screening modulators of human NuBP activity or			
CC		expression. The human NuBAPs, and their agonists and antagonists are used			
CC		to treat diseases associated with overexpression or underexpression of			
CC		functional NuBAPs. Human NuBP proteins and nucleotides, and NuBP			
CC		agonists and antagonists can be used to diagnose, treat and prevent			
CC		reproductive, immunological, neurological and cell proliferative			
CC		disorders. Reproductive disorders that may be treated using compositions			

Query Match	Best Local Similarity	6.5%;	Score 341;	DB 21;	Length 799;
Matches 160;	Conservative 66;	Mismatches 280;	Indels 214;	Gaps 23;	
278	PLGSG-HPFASGVRSHKPTPAPRALPGSTIDQLLASPHLAPFSTTGLLAQGLAARG	336			
244	PLGEKPRPCGCGGKGFST-----SPRLPLRN--VHTGCKTSOSSHLRTHQIHPEEK	295			
337	LEATASDGLLKP-----NGSGELSYEYWGPLEKDEGRHKRCRCARVFGSDSALQI	388			
296	LINCHESGDGFENKSSFHYSQSNHTGEKSY-----RCDCGKGFSSSTGLII	341			
389	HLRSHGGERPKCWCNGRRFTTGLKLYFHNRREKYPHYQMPHRYPHRIDYITSSGL	448			
342	HYRTHIEKPKCECGKCFSSQNFQCHQRYHTEKP-----YKCECGK	387			
449	PYGMVY-----PREKAEBAATPGGVEKKPLVASTALSTESTLTLSTAG	496			
388	GFGMSVNLRVHQRVHGRGKPKCEE-----	412			
497	TATAPGLPANKFLYLMKAV---EPKNKADENTPPGSEGSNISVAESSTYTLMOLSLMT	553			
413	--CGKGTQAAHFHQRVHTGKPKYKD-----YCGKGFSSNPLICHRVHT	459			
554	SLPSMAL-----LTNHF-KSTGSPFLPLCARALGASPSFSTKLOQLVEKIDRQ	600			
460	GKRPYCEACGKGTNTDLIHFRVHTGKPKY--CKCGKGFQASNLQVHONVHTGE	517			
601	GAVAVYISASGAPPTSPAPSSASSG--PNQCVTCRLVLSCPRALRLHYGOGGERPFK	658			
518	KREKCEFCGKGFSSSKLOTQRVHTGKPKYCDVCGKDFSSNKLHQVHTGKPKYK	577			
659	CKYCGRAFSTRGNLRHFVGHKASPAALAAONSCPTQCKKFTNAVITLOOHVHMLGGQIPN	718			
578	CECGGKGFMSKSNLHAQRVHSGKPKY---CEQCDKFSQIDFRVHQRVHTGKPKYK	633			
719	GGTALPEGGGAOENGSEOSTVSGASFPPOOOSOPPEELSEEEDEEEDVETDE	778			
634	CGVC--GKGFSSQSHQSHQVHTGKPKY-----K	661			
779	DSLAGSGSEBGEKALIVRGDSEASGAEFEVGYTAAAAATAGKMDNSNEKTQOSSLPP	838			
662	CDVCGKGFYSQFIYHQHHTGKPKYKCECG--KGFENST-----	701			
839	PPPSIDDPQPMKOGSSGVLAGKEGCKPERSSSPASALTTPGEATSVTLVEELSLQEAM	898			
702	----NLRHQRVHTGKPRHI--CECGK-----AFSLPSNL	731			
899	KREPESSSKK--ACEYCGAFPSSOALIEHQKTHPKRGPLTCVFCROGFLERATLTKH	956			
732	RVHLGVHTREKLFKCECGKGFSSQARLEAHQRVHTGKPKYKCDICDKDFRHSRLTYH	790			
RESULT 30					
AAH80283					
AAH80283	standard; Protein; 927 AA.				
AAH80283;					

[illegible]

QY	219	ASPSRLP-----	-----GTGASSRKRLPLPSPIKPVOTSKTLASSSSSSSSSS	263
Db	153	PAPEVVPFSDMTWKACEASWOMGALTTWNPPVPPV-----	-----ANEPISRLVQ	198
QY	264	GAEETPQGAEPHLVHPIGSOHPFSAAGVGRSHKPPAPSPALPGSDLIASPHLAFPTT		323
Db	199	G-----RPGAEPYICNEGKS-----	-----FSQMS	221
QY	324	GLIAADCLGAARLEATIASPGLLKPKNGSGLSYGEWGRPLEKPGGRHKRCRCAPVFGSD		383
Db	222	KILRHQRHITGKERPNICSEGG--KSPFOSSHLVQHQRHTHTEK P--YKCPDGCKCFWSWS		276
QY	384	SALQILRSHTGERPYKCNVGNCRFTTRGNLKVFIHRRKEXP-----		426
Db	277	SNLVQHQRHTHTEKPYKCTECEKAFQSTNLIKHQRSHTGKERPYKCGBCRRAPFYRSSDLI		336
QY	427	-----HYQNMHPVPHVLDVITSSGLRYGVSVPREKAEELAPVGGVERKPLVASTA		481
Db	337	QHQATHTGKERPYKCP E-----CGRRFGGNHLLKHKQLYHA-----GEKPYKCTEGG		382
QY	482	LSATESLTLLS---TSAGTATAPGLPAFKNR-----VLKAAEPKNKADENTPPGEGSAI		534
Db	383	KSTIOSSSLQHQRTHTGKERPYELCEGSKSGHSSTLLKHQRTHLRDRPKCP-----V		436
QY	535	SGVAESSTATLMOLSKLMTSLPSWALLTHNEKSTGSPF--PLCARALGASEFTSKL--		590
Db	437	CGKTFSLTATLLRHQR-----THTGERPYKCPDEC---GKSFVSNSNLIN		477
QY	591	QQLVERKIDRGAAVAVTSAGAPFTTSAPAPSSSASGNGCVICLRVLSCPRALRHHYQ		650
Db	478	HQRTHQER-----PYICADCKSFITMSSTLLRHORI		509
QY	651	HGRERPPKCVCGRAFSTRGNLRAHFVGHKASPARAONSCPIQCKFTNAVTLDQNHVM		710
Db	510	HTEKEPKKCDCKSFTR-----SSHLDHRRHTHTEKERYKCPCEGCKSPFOSSNLITHVRT		565
QY	711	HLGGQI---PNGSTALDEGGCAOENSGQSTVSGAGSPFOQOOSQPSPEELSEEEEEE		767
Db	566	HMDENLFVCSDCGKAFLEAHLEEQHR-----VHERGKTPPARAO-----		605
QY	768	DEEEEDVYTEDSLAGGSES-----GGEK--AISRQDSSEASGAEE-----		809
Db	606	-----GDSLILGIDPSLLTPPPAKKHKLVCCKGFNDEGIFMQRHRIHGENPY		655
QY	810	---VGTVAATAATAGKEMDSNEKTTQSSSLPPPPPSDLDQPMQEGSSGVLGKEEGK		866
Db	656	KNADGLIAHA-----PKRPOLRSPLPLPRGNS--YPGAEG--		690
QY	867	PERSSSPASALTPREGENTSVTLVEELSLQEMARKPESSSRK-ACEVCGAFPSCAAL		924
Db	691	--RAEAPGQPLKPEEG-----QEGFSQRGLISKTYICSHCGESFLDRSVL		735
QY	925	EEHQKTHPKREGPL-----FTCVFQRQGLLEATLTAKHMLLH		961
Db	736	LQHQDLTHGENRPLFLPYRIGLREGAGSPPLSGKPFCKCPCEKQSP-----GLSEILL-H		790
QY	962	HOV		964
Db	791	QKV		793
RESULT 31				
AAB95103				
ID	AAB95103 standard; Protein: 619 AA.			
AC	AAB95103:			
XX				
DT	26-JUN-2001 (first entry)			
DE	Human protein sequence SEQ ID NO:17076.			
XX				
XX	Human; primer: detection; diagnosis; antisense therapy; gene therapy.			
XX				


```
OS Homo sapiens.
XX
XX EPI074617-A2.
XX
XX 07-FEB-2001.
XX
XX 28-JUL-2000; 2000EP-0116126.
XX
XX 29-JUL-1999; 99JP-0248036.
XX 27-AUG-1999; 99JP-0300253.
XX 11-JAN-2000; 2000JP-0118776.
XX 02-MAY-2000; 2000JP-0183767.
XX 09-JUN-2000; 2000JP-0241899.
XX
XX (HELI-) HELIX RES INST.
XX
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
XX Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX WPI; 2001-318749/34.
XX
XX Primer sets for synthesizing polynucleotides, particularly the 5602
XX full-length cDNAs defined in the specification, and for the detection
XX and/or diagnosis of the abnormality of the proteins encoded by the
XX full-length cDNAs -
XX
XX Claim 8: SEQ ID 17076; 2537pp + CD ROM; English.
XX
XX The present invention describes primer sets for synthesizing 5602
XX full-length cDNAs defined in the specification. Where a primer set
XX comprises: (a) an oligo-dT primer and an oligonucleotide complementary
XX to the complementary strand of a polynucleotide which comprises one of
XX the 5602 nucleotide sequences defined in the specification, where the
XX oligonucleotide comprises at least 15 nucleotides; or (b) a combination
XX of an oligonucleotide comprising a sequence complementary to the
XX complementary strand of a polynucleotide which comprises a 5'-end
XX sequence and an oligonucleotide comprising a sequence complementary to a
XX polynucleotide which comprises a 3'-end sequence, where the
XX oligonucleotide comprises at least 15 nucleotides and the combination of
XX the 5'-end sequence/3'-end sequence is selected from those defined in
XX the specification. The primer sets can be used in antisense therapy and
XX in gene therapy. The primers are useful for synthesizing polynucleotides,
XX particularly full-length cDNAs. The primers are also useful for the
XX detection and/or diagnosis of the abnormality of the proteins encoded by
XX the full-length cDNAs. The primers allow obtaining of the full-length
XX cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
XX AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
XX AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
XX represent oligonucleotides, all of which are used in the exemplification
XX of the present invention.
XX
XX Sequence 619 AA:
XX
XX Query Match 6.38; Score 333.5; DB 22; Length 619;
XX Best Local Similarity 21.08; Pred. No. 3.8e-11;
XX Matches 170; Conservative 83; Mismatches 296; Indels 259; Gaps 30;
XX
XX 217 ADASPSELPGTASTTKPLPLPSPIKPVQTSKTLASSSSSSSSSGAETPKQAFHNL 276
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX 34 APRARNAHTNGRADSEYV---AEVKRPETTEKAEBAASEKYSGSAKRPYACRPLC 89
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX 277 HPLGSONPFSAGVGNSH---KTPRAPSPALPGSTDLIASP-----HLAFSTTGILAA 328
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX 90 -PKAYKTAPELRSHSGSHTEKRPSP-----ECGRRTMQVCLKVHLA----- 132
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX 329 QCLGAARGLDAPASPLGLKKNKSGELSYGEVMGPLEKRGGNKCFKAKVGSDSALQI 388
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX 133 -----SHAGELPF-----RCANHCPRKAYGALSCLKI 157
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX 389 HLKSHGERPYKCNVCGNRRFTTNGNLKVHFRHREKPYHVOVMPHVPRLDVLITSSGL 448
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX 158 HQRGHTGERPYACADCGSKPADPS-----VFKKHRR--THAGLRP-----YSCERCGK 203
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | |
```

```
QY 449 PYGMSVPEPEKAEEAATPGGVERKPLVAST--TALSTATESLTLTSTAGTATAPGLPAF 506
XX | : : : : : : : : : : : : : : : : : : : : : : : :
DB 204 AYA-ELKDLRHNHERSHT-----GERPFLCSECGKSFSSSSITC HORIHAOKPYRCAC 257
XX | : : : : : : : : : : : : : : : : : : : : : : : :
QY 507 NK-FVLKAVPEPKNAKADENTPPESEGSATISGVASSTATIMQLSKMLTSLPSMALITNHF 565
XX | : : : : : : : : : : : : : : : : : : : : : : : :
DB 258 GKGFYQLSSYSHERTHSGEK----- 279
XX | : : : : : : : : : : : : : : : : : : : : : : : :
QY 566 KSTGSPFLPLCARALGASPSSETSKLQQLVEKID-----RQCAVAVTSAGCAPT 614
XX | : : : : : : : : : : : : : : : : : : : : : : : :
DB 280 -----FLCPRCGR-MFSQPSFRRHQRAHEGVKPYHCEKCKDFRQ-----PA 321
XX | : : : : : : : : : : : : : : : : : : : : : : : :
QY 615 TSAPAPSSSASSGPNQVTCILVLSQPRALRLHYGQHGGERPPEPKCYGCAFTSRGNARA 674
XX | : : : : : : : : : : : : : : : : : : : : : : : :
DB 322 DLAMHRRVHTGDRPFKLCQCDTETVYASWDLKRLALVHSGQRPCEGCGRAFAERASLTK 381
XX | : : : : : : : : : : : : : : : : : : : : : : : :
QY 675 HFVGHKASPAARAQNSCPICQKFTNAVTLQOHVNMILGQ---IPRG-----GTALPEG 726
XX | : : : : : : : : : : : : : : : : : : : : : : : :
DB 382 HSRVH-----SGERPFHCNACGKSFYVSSSLRKHERTHRSSEAGAVPPAQLVGLALPY- 436
XX | : : : : : : : : : : : : : : : : : : : : : : : :
QY 727 GGAQENGSEOSTVSGAGSFPQOQSQSPSEBELSEEEDEDEEEDVDEDSLGRGS 786
XX | : : : : : : : : : : : : : : : : : : : : : : : :
DB 437 -GVAQESSAPAPAGAGLQDPPA-----GILGLPP 464
XX | : : : : : : : : : : : : : : : : : : : : : : : :
QY 787 ESGGEKA-----ISVRDSEEAAGAEVEGVAAATAAGKEMDSNEKTTQSSLP 839
XX | | | | | : : : : : : : : : : : : : : : : : : :
DB 465 ESGGVMAITQKQVGNKTVAEHVECDAGVRAAPRLPEGAGEAGE-EADEK----- 512
XX | | | | | : : : : : : : : : : : : : : : : : : :
QY 840 PPSLDQPPPEGSSGVLGKKEGKPERSSSPASALTPEGEATSVTLVEELSLQEAMR 899
XX | : : : : : : : : : : : : : : : : : : : : : : : :
DB 513 PPOFVCH-ECKETSTMTLLRHNERSHPELRFP----- 545
XX | : : : : : : : : : : : : : : : : : : : : : : : :
QY 900 KEGGESSRKACEVCGGAFFQQAALAEHQTKPKESPLFTVCVRCGFLERATLKKHML 959
XX | : : : : : : : : : : : : : : : : : : : : : : : :
DB 546 -----CTQCQSISPDAGLRKHSRTSSVAP-YTCHCPKPAFLASDLRKEH-- 591
XX | : : : : : : : : : : : : : : : : : : : : : : : :
QY 960 AHHQVQPFAPHGPNQIALSLVPGCSPS 987
XX | : : : : : : : : : : : : : : : : : : : : : : : :
DB 592 RTHPVPGTPTPLPPIVALLQMPREGPA 619
XX | : : : : : : : : : : : : : : : : : : : : : : : :
RESULT 32
AAB94669
ID AAB94669 standard; Protein; 498 AA.
XX
XX AAB94669;
XX
XX 26-JUN-2001 (first entry)
XX
XX Human protein sequence SEQ ID NO:15609.
XX
XX Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX
XX OS Homo sapiens.
XX
XX EPI074617-A2.
XX
XX 07-FEB-2001.
XX
XX 28-JUL-2000; 2000EP-0116126.
XX
XX 29-JUL-1999; 99JP-0248036.
XX 27-AUG-1999; 99JP-0300253.
XX 11-JAN-2000; 2000JP-0118776.
XX 02-MAY-2000; 2000JP-0183767.
XX 09-JUN-2000; 2000JP-0241899.
XX
XX (HELI-) HELIX RES INST.
XX
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
XX Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX WPI; 2001-318749/34.
```


XX primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
PS Claim 8; SEQ ID 15609; 2537pp + CD ROM; English.
XX
CC The present invention describes primer sets for synthesizing 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX
XX Sequence 498 AA:
Query Match 6.2%; Score 329.5; DB 22; Length 498;
Best Local Similarity 22.0%; Pred. No. 5e-11;
Matches 146; Conservative 70; Mismatches 240; Indels 209; Gaps 24;
OY 352 SGEIATYGVWVGRPLKPGGRHRCFCAGKAVGSDSALQIHLRSHGERYKVCNCRFTTR 411
DB 14 AGELPF-----RCANCPKAGALSKIKIHQGHGGERYACADCKSKSPFDP 59
OY 412 GNLKHFHRHREKYRVQMNPRVPEHLDYVITSSGLPYGMSVPERKAEERAPGGGVE 471
DB 60 S-----VERKHNRR--THAGLRP-----YSCERCGKAVN-ELKDLRNNHESNR-----G 99
OY 472 RKPLVAST--TALSATESLTLLSTAGTATAPGLPAPFNK-FVLKKAVERPKKADENTPRG 528
DB 100 ERPEFLCECGKSFSSSSSLTCHQRTIHAAQKPYRCACGKGTOLSSYQSHERTISGEKP- 158
OY 529 SEGSAISGVASSSTATLMQLSKMTSLPSWALLTNHFKSGSPFLPCARLGLGASPSPTS 588
DB 159 -----FLLCPRCGR-MFSDPSFPR 175
OY 589 KIQQLVLEKID-----RQGAVAVNTSAASGAPTTSPADSSASSGPNOCTICLRV 637
DB 176 RHQAHGKGVKRYHCKGCKDRFO-----PADLAMHRRVNHGDRFKICQDCOKT 223
OY 638 LSCPRALRLHYGNGGERPFCKVCVGRAFSTRGNLRAHFGVGNKASPARAQNSCPICOKK 697
DB 224 FVASWMDLKRHLAIVHSGORPFCRCECGRAFAERASLTGHSRVH---SGERPFHCNACGKS 279
OY 698 FTNVAVTLOOHVBMHNGGQ---IPNG-----GTALPEGGGAOENGSEGSTVSGAGSPFOQ 749
DB 280 FVSVSSLLKHKHRTTHSSSEAGVPPPAOELVVGALPV--GVAEGSSAPARACAGLGDPPR- 336
OY 750 QSQQPSPEELSEEEEDDEEEDVTDSDSLAGRSGSEGEKA-----ISVRGDSBE 802
DB 337 -----GLGLRPESSGVMTQMGVGVGTVAHVHQ 366
OY 803 ASGAEEEVGTAAATAAGKEMDSNEKTTQOOSLPPPPPDLDPQPMEGSSGVYLGEKE 862
DB 367 DAGVREAPGLEGAGEAGE-BADEK-----PQFVCR-ECKETSTMTLLRRH 413

OY 863 EGKPERSSPASALTPEGEATSVTLVEELSLQEMARKPEGESSRKAACEGCAFPSSQA 922
DB 414 ERSHPELRPPP-----CTQCGKSFSDRA 436
OY 923 ALBEOKHTHPKEGFLTCVFCRQGLERATLTKKMLAHNQVOPFAPHGPNIALSLVP 982
DB 437 GLRKHSRTHSSVPR-YTCRPHCKAFLSASDLRKHE--RTHRPVPMGRPTLEPLVALLGMP 493
OY 983 GCSPS 987
DB 494 EEGPA 498
RESULT 33
AAB95116
ID AAB95116 standard; Protein; 725 AA.
XX
XX AAB95116:
AC 26-JUN-2001 (first entry)
XX
XX
DE Human protein sequence SEQ ID NO:17104.
XX
XX Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX Homo sapiens.
XX
XX EP1074617-A2.
XX
XX 07-FEB-2001.
XX
XX 28-JUL-2000; 2000EP-0116126.
XX
XX 29-JUL-1999; 99JP-0248036.
XX 27-AUG-1999; 99JP-0300253.
XX 11-JAN-2000; 2000JP-0118776.
XX 02-MAY-2000; 2000JP-0183767.
XX 09-JUN-2000; 2000JP-0241899.
XX
XX (HELI-) HELIX RES INST.
XX
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
XX Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX WPI; 2001-318749/34.
XX
XX Primer sets for synthesizing polynucleotides, particularly the 5602
XX full-length cDNAs defined in the specification, and for the detection
XX and/or diagnosis of the abnormality of the proteins encoded by the
XX full-length cDNAs -
XX
XX Claim 8; SEQ ID 17104; 2537pp + CD ROM; English.
XX
XX The present invention describes primer sets for synthesizing 5602
XX full-length cDNAs defined in the specification. Where a primer set
XX comprises: (a) an oligo-dT primer and an oligonucleotide complementary
XX to the complementary strand of a polynucleotide which comprises one of
XX the 5602 nucleotide sequences defined in the specification, where the
XX oligonucleotide comprises at least 15 nucleotides; or (b) a combination
XX of an oligonucleotide comprising a sequence complementary to the
XX complementary strand of a polynucleotide which comprises a 5'-end
XX sequence and an oligonucleotide comprising a sequence complementary to a
XX polynucleotide which comprises a 3'-end sequence, where the
XX oligonucleotide comprises at least 15 nucleotides and the combination of
XX the 5'-end sequence/3'-end sequence is selected from those defined in
XX the specification. The primer sets can be used in antisense therapy and
XX in gene therapy. The primers are useful for synthesizing polynucleotides,
XX particularly full-length cDNAs. The primers are also useful for the
XX detection and/or diagnosis of the abnormality of the proteins encoded by
XX the full-length cDNAs. The primers allow obtaining of the full-length
XX cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
XX AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
XX AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632


```

Db      61 MSLEKSSDDPSVQNPPTP---LGHSN-PL-----DHQIPD----- 93
Oy      210 SLGQTVGAPASPSLP-----GTGTASTKPLPLFSPDKPYQTSKTLAS 254
Db      94 -----PPAPEVVPPTPSDWTKACASMQMGALTTWNSPPVVP-----AN 131
Oy      255 SSSSSSSSGAETPKQAFHLHLNPLCSQHPPFSAGVGRSHKPTAPSPALPGSTDLIAS 314
Db      132 EPSLRELVOG-----RPAGAEKPYICNECKS----- 158
Oy      315 PHLAPPTSTGLLAACIGAAAGLEATASPGLLKPKNGSGELSYGEVMPLKPGGRHKCR 374
Db      159 ----FSQMSKLLRHQRIHGERPNTSECG--KSFQSSHLVQHRTTGERP---YKCP 209
Oy      375 FCAYVFGSDSALQIHLSRHTGERPYKCNVGNRETTGNLKHVHRHREKYP----- 426
Db      210 DCGKCFSSSNLVQHQRTHTGERPYKCTECEIAFTQSTNLKHQSHHTGERPYKCGECRR 269
Oy      427 -----HYQMNPHVPENLDVYITSSGLPYGMVPPPKAEAEATPGGVER 472
Db      270 AFYRSSDLIQHQAHTHGERPYKCP-----CGKRGQHNLKHKQKIH-----GE 315
Oy      473 KPLVASTALSTESTLTS---TSAGTATAPGLPAFNKF---VLKMAVEPKNKADENT 525
Db      316 KPYRCTEGCKSFQSSSELTOHQRTHTGERPECELECGKSFQHSSTLIKHQRTHLRDPER 375
Oy      526 PPGSEGAISGVAESSTATLMQLSKLMTSLPSMALLTNHFKSTGSPF--LPLCARALGAS 583
Db      376 CP-----VCGKTFTLSTATLLRHQR-----THTGERPYKCP-----GKS 410
Oy      584 PSETSKL--QQLVERKIDRQCAVAVTSAASCAPTSAAPASSASSGPNQCVICRLSCP 641
Db      411 FSVSSNLIHQRIHREGR-----PYCACGCKSFIMS 442
Oy      642 RALRLHYGQHGGERPKCKVCYCGRAFTRGNLRAHFVGHKASPARAQNQSCPIQCKFTNA 701
Db      443 STLIRHQRIHTGKPKYKSCGKSFIR---SSHLIQRIHTGKPKYKCEGCKSFQS 498
Oy      702 VTLOQHVNMHLGQI---PNGGIALPREGGAQENGSEOSTVSGAGSPQOQSOPSPPEE 758
Db      499 SNLITVRIHTHMDNLFVCSDCGKAFLNAHLEQHR---VIHERKKTPARRAO----- 547
Oy      759 ELSEEEEDDEEEDVTDDSLAGRSES-----GGEK---AISVRDSEFASGAEE 809
Db      548 -----GDSLIGLGDPSLLTTPPGAKPKKCLVCGKGFNDGCTFMQHQ 588
Oy      810 -----VGTVAATAAGKEMDSNEKTTQOSSLPPPPPSLDOPQPMEOGSSGV 857
Db      589 RIHIGENPYKNAAGLIAHNA-----PKPQLRSPRLPFRGNS-- 625
Oy      858 LCGKEGCKPERSSSPASAL--TPREGATSVTLVEFLSLQEMARKKEGESSRK--ACVCG 915
Db      626 YPGAAGC---RAEAPQCPKPRPG-----OEGFSQIRGLLSSKITYICHG 668
Oy      916 QAPPSQALEEHOKTTHKEGPL-----FTCYFCROGFLERAT 952
Db      669 ESTLDRSVLQHQHTHGENEKFLLPDRIRIGLEGAGPPFLSGKPRKCECKQSF-----G 724
Oy      953 LKKHMLLAHHQV 964
Db      725 LSELELL-HQKV 735

RESULT 35
ABG06465
ID      ABC06465 standard; Protein: 882 AA.
XX
AC      ABG06465;
XX
DT      13-FEB-2002 (first entry)
XX
DE      Novel human diagnostic protein #6456.
```

```

XX      Human: chromosome mapping; gene mapping; gene therapy; forensic;
KW      food supplement; medical imaging; diagnostic; genetic disorder.
XX      Homo sapiens.
XX      WO200175067-A2.
XX      11-OCT-2001.
XX      30-MAR-2001; 2001WO-US08631.
XX      31-MAR-2000; 2000US-0540217.
XX      23-AUG-2000; 2000US-0649167.
XX      (HXSE-) HXSEQ INC.
XX      Drmanac RT, Liu C, Tang YT;
XX      WPI: 2001-639362/73.
XX      N-PSDB: AAS70652.
XX      New isolated polynucleotide and encoded polypeptides, useful in
XX      diagnostics, forensics, gene mapping, identification of mutations
XX      responsible for genetic disorders or other traits and to assess
XX      biodiversity.
XX      Claim 20; SEQ ID No 36824; 103pp; English.
XX      The invention relates to isolated polynucleotide (I) and
XX      polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX      polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX      and gene mapping, and in recombinant production of (II). The
XX      polynucleotides are also used in diagnostics as expressed sequence tags
XX      for identifying expressed genes. (I) is useful in gene therapy techniques
XX      to restore normal activity of (II) or to treat disease states involving
XX      (II). (II) is useful for generating antibodies against it, detecting or
XX      quantitating a polypeptide in tissue, as molecular weight markers and as
XX      a food supplement (II) and its binding partners are useful in medical
XX      imaging of sites expressing (II). (I) and (II) are useful for treating
XX      disorders involving aberrant protein expression or biological activity.
XX      The polypeptides and polynucleotide sequences have applications in
XX      diagnostics, forensics, gene mapping, identification of mutations
XX      responsible for genetic disorders or other traits to assess biodiversity
XX      and to produce other types of data and products dependent on DNA and
XX      amino acid sequences. ABG00010-ABG30377 represent novel human
XX      CC diagnostic amino acid sequences of the invention.
XX      CC Note: The sequence data for this patent did not appear in the printed
XX      CC specification, but was obtained in electronic format directly from WIPO
XX      CC at ftp.wipo.int/pub/published_pcl_sequences.
XX      SQ      Sequence      882 AA;
XX
XX      Query Match      5.9%; Score 311.5; DB 22; Length 882;
XX      Best Local Similarity 21.9%; Pred. No. 1.1e-09;
XX      Matches 180; Conservative 82; Mismatches 304; Indels 255; Gaps 34;
Oy      18 PABLGASERDHDQVCAKCCAGFTDPTEFLAHQNCSTDPVMIITGGENPNSSASS 77
Db      129 PLOVLGSHKEKENLKP-AKAOLPFTTGPFTTNDREKESP-----SSTWNPSEMQRKA 181
Oy      78 EPRPEGHN-----NPQVADTEHSNPPDSGSSVPTDPTWGPERRGESSGHFLVAAT 128
Db      182 SPRKWKHHNQRPSTHKVKKMWSSEOMKLPSTKKAEP--PTWQOLKK----- 226
Oy      129 GTAAGGGGGLLASPKGATPLPPESTPAAPPPPPPPPGVSGHLNIPLILEKVLQ 188
Db      227 -----LTOLPKKS-----LENTKV-- 240
Oy      189 QROTHQMQMTQICROVLLIGSLGQTVGAPASPSLPQGTASTKPLPLFPSP-----K 244
Db      241 -----TOTPENTLIAALMIVS---TAGAAAA-----NTTWAYVYVFPPLIRAVTMDN 285
```


OY	245	PVQGSKTLASSSSSSSSSGAEITPQAQFHLHYLPGLGSDHPFSAGVGNSHKPTPADSPAL	304
Db	286	PIEIVYNNASAVWPPETDRCDPKDEEGEMMINISTGYRPPIC--LGR-----APACIM	337
OY	305	PGSTDOLIASPHLA----FPSTTGLIAA-----OCIGAA-----RGLEATA	341
Db	338	PAIONMLVEVPVTGTSTFTHTTGILTFTRDAVIERFSQEWMKKLDPALEDSTYRRKANSGC	397
OY	342	SP-----GLIKPRNG-----SGE-----LSYGEVMGPLEKPGG	369
Db	398	KPKPCNECGAFNTSNSLTSHRHHSIGSEKPYKCNECGAKTFVRSNLTTHQVIHTEGKP--	455
OY	370	RHKRCFAKYFGSDSALOIHRSHTGEERPYCNVCGNFFTTNGTKVHHHRREKYP----	426
Db	456	-YNCHCEGGAYFRHNYSYLATHRIHTEGKEPYKCNECGAKFARGHNSLTTHQLIHTEGKEPK	514
OY	427	-----HVNMPHPPEHLVDYVTSSTGLEYGMNV-----PPEKAEE--	461
Db	515	NECGKLFTONSHLISHWRITHTEGKEPYKCNECGAKFVSRSLSIAHQIHTGEKPYKCNEGC	574
OY	462	EAAIPGGGVSRKPLVASTATLASATESLLTSTAGTARGLPARNKVYMKAVEREKNA	521
Db	575	KYRVNGVGS-----IGTCCLISVFQVITVSSRK-----RASDENRA	612
OY	522	DEN---TPPSGEGSAISGAESSTATLMQLSKMTSLPSWALLTNHFKSTGPSF--LPL	575
Db	613	TQSILLAPPTSRSQG--PGPA-SVLRLIAQRKRIR---AW---TESCTAACAPSCVL	663
OY	576	CATALGASPE---TSKLOQLVEKIDROGANAAVAISAAGCAPTTSAPARSSASSG-----	627
Db	664	LYNGRGV-PTDLOYPAFLFSTQSKLKRPGSTKSSLATKHORIHTGEKPYKCNECGKVFSQ	722
OY	628	-----PNQCIVLCRLVLSCPRALRLHYGOHGGERPFCKVCGRAFSTRGNL	672
Db	723	TSSLARHWRIHTEGKEPYKCNECGKVFVSYNSHLASHRYVHTGEKPYKCNECGKAFSVHSNL	782
OY	673	RAHFVGHKASPAPAQNSCPICQKKFTFNNAVTLQOHVRMHLG	713
Db	783	THQVIHTEGKEPYK---CNCGKGKGFVSHSSLTHQVIHTEG	819
RESULT	36		
ABB69958	ID	ABB69958 standard; protein; 934 AA.	
XX	XX	ABB69958;	
XX	XX		
DT	26-MAR-2002	(first entry)	
XX	XX		
DE	Drosophila melanogaster polypeptide SEQ ID NO 36666.		
XX	XX		
KM	Drosophila; developmental biology; cell signalling; insecticide; pharmaceutical.		
XX	XX		
OS	Drosophila melanogaster.		
XX	XX		
PN	MO200171042-A2.		
PD	27-SEP-2001.		
XX	XX		
EF	23-MAR-2001; 2001WO-US09231.		
XX	XX		
PR	23-MAR-2000; 2000US-191637P.		
PR	11-JUL-2000; 2000US-0614150.		
XX	XX		
PA	(PEKE) PE CORP NY.		
X1			
P1	Venter JC, Adams M, Li PMD, Myers EW;		
XX	WI: 2001-656860/75.		
DR	N-PSDB; ABL14061.		
XX	XX		
TF	New isolated nucleic acid detection reagent for detecting 1000 or more		

[illegible]

QY	824	DSMEKTTQSSLPPEPPDSDLP---QPMQGS--GVLGKKEGGK--PERSSP---	873
Db	1311	-SPPPKKSNSPANPVATTTWQEPATAPSTNPSPSLKMTIMQAEVVGSLKKEVASSPFKD	1369
QY	874	-ASALTPEGE---ATSVT-----LVELSLQE-----	896
Db	1370	ESQDLPVAVATLVNATQNNQNGFNFSFRPSDVANHEQSDDEGLVASSGASSENNGTEDI	1429
QY	897	---AMKREPESS-----SRKACEVCGQAFPSQAALAEHOKTHPKESPLFTVCRCQGF	947
Db	1430	TSSSSSEPKKSAVSLAPNRVSCPYCQRMFPWSSSLRRHILHTTGQKP-FKCSHCPLLF	1488
QY	948	LERATLTKHMLIAHMOYQ 965	
Db	1489	TTKSNCDRHILRRKHGNYE 1506	
RESULT 38			
AAM39029			
ID	AAM39029	standard; Protein; 722 AA.	
AC	AAM39029;		
XX	22-OCT-2001	(first entry)	
DE	Human polypeptide SEQ ID NO 2174.		
XX	Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;		
XX	peripheral nervous system; neuropathy; central nervous system; CNS;		
XX	Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;		
XX	amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;		
XX	chemokine; thrombolytic; drug screening; arthritis; inflammation;		
XX	leukaemia.		
OS	Homo sapiens.		
XX	WO200153312-A1.		
PD	26-JUL-2001.		
XX	26-DEC-2000; 2000WO-US34263.		
PF	21-JAN-2000; 2000US-0488725.		
XX	25-APR-2000; 2000US-0552317.		
XX	09-JUL-2000; 2000US-0598042.		
PR	19-JUL-2000; 2000US-0620312.		
PR	03-AUG-2000; 2000US-0653450.		
PR	14-SEP-2000; 2000US-0662191.		
PR	19-OCT-2000; 2000US-0693036.		
PR	29-NOV-2000; 2000US-0727344.		
XX	(HYSE-) HYSEQ INC.		
PA			
XX	Tang YF, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;		
PI	Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;		
PI	Zhao QH, Zhou P, Goodrich R, Demanac RT;		
XX	WPI: 2001-442253/47.		
DR	N-PSDB; AAI58185.		
XX			
PT	Novel nucleic acids and polypeptides, useful for treating disorders		
XX	such as central nervous system injuries -		
XX	Example 4; SEQ ID NO 2174; 10078pp; English.		
PS			
CC	The invention relates to human nucleic acids (AA157798-AA161369) and		
CC	the encoded polypeptides (AAM38642-AAM42213) with nootropic,		
CC	immunosuppressant and cytostatic activity. The polynucleotides are useful		
CC	in gene therapy. A composition containing a polypeptide or polynucleotide		
CC	of the invention may be used to treat diseases of the peripheral nervous		
CC	system, such as peripheral nervous injuries, peripheral neuropathy and		
CC	localised neuropathies and central nervous system diseases, such as		
CC	Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic		

CC	lateral sclerosis, and Shy-Dreager Syndrome. Other uses include the
CC	utilisation of the activities such as: Immune system suppression,
CC	activation/inhibition activity, chemotactic/chemokinetic activity, haemostatic
CC	and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC	assays for receptor activity, arthritis and inflammation, leukemias and
CC	C.N.S disorders.
CC	Note: The sequence data for this patent did not form part of the printed
CC	specification.
xx	
xx	Sequence 722 AA;
xx	Query Match 5.7%; Score 303; DB 22; Length 722;
xx	Best Local Similarity 22.4%; Pred. No. 2.6e-09;
xx	Matches 164; Conservative 70; Mismatches 271; Indels 228; Gaps 27;
QY	348 PKNGSGELSYGEVWGPLEKPGGRH-KCRFCAYFGSDSLQILHLSHTGTERPYKNCGN 406
DB	33 PAYSAGSLGMGAVSMSESERAGERRRPPCYCGRRFNSTLALHRAHGAQAFPCPHGH 92
QY	407 RFTTRGMLKAFHFNHREKYPRHYQAMPHPVEHLIDYITSSGLPYGMSVPEKAEEDATP 466
DB	93 RAAQALLRSLTLTHQRRP----- 112
QY	467 GGGVERPLVASTALSTESLTLLSTAGTAPGLPAFNKFLVLMKAVBPKNKADENTP 526
DB	113 -----RSP--AARILLEERALLREARLGRARSSG-----GQAT 146
QY	527 PGSEGSASISGVAESSTATLMOLSKLMTSLPSWALLTNHFKSTGSFPLPLCARLAGASPE 586
DB	147 PATEGLARPOAPSS-----AFRCPYCKGFRFS-AE 177
QY	587 TSKIOQLVERKIDRGAVAVTSA-----ASGAP-----TTSAPAPSSASSGPN 629
DB	178 REKHHLTLHPRWKCGKGLSGSSQEBELHLSLTAGCAPERPLAATSAAPROPOROPPO 237
QY	630 QCVICLVLSCPRALRLHYGGHGER-----PRCKVCGRAFTSGN 671
DB	238 -----PEPRVPQREPEREREATPTPAPAREEPAPPEPRQCQGSFTQSWF 288
QY	672 LRAHFGHKASPARAANSCPTIOCKKFTNATVLOOHVKN-----LGGIIPNGTAL-- 723
DB	289 LKGHMKRHKAS-----FDHACPVCGRCFEPWFLKNHMKHASKLGLRLAPGAPSGPARAP 344
QY	724 --PEGGAQENGSEOSTVSGAGSFPOQSOOPSSEELSEEEDEDEDEEDVTEDSL 781
DB	345 QRPDLGLLAEYELGRLALLAPAT-PAERREPPSLGLYSLRAGEGR-----NGEGAE 397
QY	782 AGNGSESG-----EKALSVGRDSEASAAEEVYVAAAA---TAGKENDSNKTTQSS 834
DB	398 PGGRGFRGGPRLLSSALPARARRHRAEPEEEVEAEQETWAKRSLSG-----LAS 451
QY	835 LPPRPDSDIDOPMEOGSSGVLGKKEEGGKPERSSPSALATPREGATSVTLVEFLS- 893
DB	452 LHRPR-----GGGPGHSASAGA-----QARSTQIENG 482
QY	894 LQEAHKEPESESSRKACEVCGAPFPGALAEENOKTHPKESPLFTCVPRQGFLEBATL 953
DB	483 LVGGTPEEGRGATGKCPCFGKSFRSNHLKVLHRLVHNGEER-XYCPHCDYAGTOSGL 541
QY	954 KKHMLLHNVQ-----PFAP-----HGPNIALSLVP-----GCSPTSTGT 992
DB	542 KHN-LQHNHREQSGAGRGPPRPSPQKASAPQSGAKPSPQATWBGASSPBPSSG 600
QY	993 LSPFPRKDDPTIP 1005
DB	601 AGPGSRR-KPASP 612
RESULT 39	
ABB59449	
ABB59449	standard; Protein; 744 AA.
ABB59449;	

XX 26-MAR-2002 (first entry)
DT Drosophila melanogaster polypeptide SEQ ID NO 5139.
DE Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX Drosophila melanogaster.
OS WO200171042-A2.
PN 27-SEP-2001.
PD 23-MAR-2001; 2001WO-US09231.
XX 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX (PEKE) PE CORP NY.
PA Venter JC, Adams M, Li PWD, Myers EW;
PI MPI; 2001-656860/75.
DR N-PDB; ABL03552.
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PS interactions -
PS Disclosure; SEQ ID NO 5139; 21pp + Sequence Listing; English.
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA
CC sequences (AB101840-AB16175) and the encoded proteins
CC (AB857737-AB872072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/Pub/Publicated_pct_sequences.
XX
SQ Sequence 744 AA:
Query Match 5.7%; Score 298.5; DB 22; Length 744;
Best Local Similarity 17.1%; Pred. No. 4.9e-09;
Matches 174; Conservative 88; Mismatches 325; Indels 429; Gaps 24;
QY 90 MDTEHSNPDDSGSVPTDPTWGPERGESESSGHFLVATGTAAGGGGILLASPKG--- 146
DB 1 MPTSESSSEITSGGGGATPMLRPSRMDQFMNS---MAAAAAAGGGGLPGADRNGGSG 56
QY 147 -----APPLPPESTRP-----APPPPPPP----- 164
DB 57 GSDGSGONGNGSRNSSASRIASVENQLAVQOHLAGLHGRPPRRPSHHREISAFVPLV 116
QY 165 ---PPPPGVSGHLNPLLELAEFLVLOOQRIHQOMTEQICRQVLLGSGIVGAPASPS 222
DB 117 TGVKVRPGSNSNEYIIMAMDKRRELALRE-----AAAAAAMGRRPGCGGGRG 164
QY 223 ELPGGTASTRPLPLFLSPIKIVQTSKTLASSSSSSSSSGAETPRKQAFHLYHPLGSG 282
DB 165 -VPPPGV-----LYGPAQV 178
QY 283 HFFSAGVGRSHKTPAPSPALPGSTDQLIASPHLAPSTTGLAQCQ--GAARGLAET 340
DB 179 PEPVYLWG-----PGSPITGAS-----PFPFGAAAAALPRPGIGRMHAG 219
QY 341 ASFGLLKPKNGSGELSYGEVMPLEKPGGRHKRCRCAKVFGSDSLQTHLRSHTGEPYK 400
DB 220 LDRRLLR-----APGRASRPKKQFICKFCNROFTKRSYNLLIHERHTTDERPYS 267

QY 401 CNVCGNRFETTRGNLKVHFNRHREKYPHVQNMHPVPEHLDVITSSGLPYGNSVPERKAE 460
DB 268 CDICKAFRRQDHLDRKRIHSEKRP----- 293
QY 461 EEATPPGGVERKPLVASTTALSATESLLTSTAGTATAPGLPAFNKFLMKAVEPKK 520
DB 294 ----- 293
QY 521 ADEMTPPGSEGAISGVAESSTATLMQSLKMTSLPSMALLTNHFKSGSPPLPCARAL 580
DB 294 -----FK----- 295
QY 581 GASPESTSKLQQLVEKIDRGAAVATSAAGAPTTSAFAPSSASSGPNOCVICTRLVISC 640
DB 296 -----CTEGCKGFGCQ 305
QY 641 PRALRLHYGQHGGERPFKCKVCGRAFSTRGNLRAHFYGNKASPAARAONSCPICQKFTN 700
DB 306 SRTLAVNHKILHMEESPDKCPVCSRFNORSLNKTHTLTH---FDHKPEYECSSCGKVFRR 361
QY 701 AVTLQOHVAMHLGGQILPNCSTA-----LPEGGG 728
DB 362 NCDLRNHALTHAVGEVNSGDYVDVEEDEARNLSGDEEDSLLEVDSPROSPVHNLGEGG 421
QY 729 AAOENGSEOSTVSGAGSFPOQSOQP-----SPEELSE-----EEDEDEEEDVYT 776
DB 422 GSEKSESEMRKLKRAALIDHEESEEDFDDDEELQDLPRVNDLPREDDDFRDEDEQ 481
QY 777 DEDSLAGR-----GSESGEKAIVRGDSEASGAEEVGTVAATAAGKEMDSNEKTT 830
DB 482 AEVALVARQASKAANTSOSSSVGTKPEROGVTHCHNEGETYMRPHGEKHQEEPGNS 541
QY 831 QOSSLPFRP-----PPBSLDQRPQMEGSSGVLGKEBEGKPRRSSPA--SALTREG 881
DB 542 GIASLPVPSPSVKYSVPPGAAGRRP-----APGAPRTNQHGNHILPENG 589
QY 882 EATSVTLVELSLQEAMRKPKGESSSRKACEVSCQAFPSQAALBEHOKTNPKEGRLTCV 941
DB 590 D-----PYLRLVHNRDLHNKSLNLSKAGVPPRRPHRTIITQPESGKPPNQRLHSPH 642
QY 942 FCRQGLERATLKKHMLANHOVQFRAFHGFPONTAALSFLVPGCSPTSTGSLSPF 997
DB 643 EAMPSEFGSIPMKRKLIPA--PTIDLMDRHNHRGIGQRTFVD--SPSIVALNMSRNP 695
RESULT 40
AB866002
ID AB866002 standard; Protein; 744 AA.
XX
AC AB866002;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 24798.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX Drosophila melanogaster.
OS WO200171042-A2.
PN 27-SEP-2001.
PD 23-MAR-2001; 2001WO-US09231.
XX 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX (PEKE) PE CORP NY.
PA Venter JC, Adams M, Li PWD, Myers EW;
PI

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 13, 2003, 15:17:45 ; Search time 15.0224 Seconds
(without alignments)
1968.393 Million cell updates/sec

Title: US-09-988-117-1

Perfect score: 5277
Sequence: 1 MAHESERSRLGVPAGEPAE.....PSTRTGSLSPPKDDPTIP 1005

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database : Issued Patents, AA:*
1: /cgn2_6/ptodata/1/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/1/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/1/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/1/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/1/1aa/PCtUS.COMB.pep:*
6: /cgn2_6/ptodata/1/1aa/backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	289.5	5.5	1185	4	US-09-041-886-23 Sequence 23, Appl
2	246.5	4.7	727	2	US-08-475-844-9 Sequence 9, Appl
3	246.5	4.7	727	5	PCT-US95-08429-9 Sequence 9, Appl
4	246	4.7	543	4	US-09-362-123A-4 Sequence 4, Appl
5	242	4.6	803	4	US-09-063-035-2 Sequence 2, Appl
6	241	4.6	728	2	US-08-475-844-5 Sequence 5, Appl
7	241	4.6	728	5	PCT-US95-08429-5 Sequence 5, Appl
8	238.5	4.5	1706	2	US-08-459-568-2 Sequence 2, Appl
9	238.5	4.5	1706	3	US-08-399-411-2 Sequence 2, Appl
10	238.5	4.5	1706	2	US-08-516-859A-2 Sequence 2, Appl
11	238.5	4.5	1706	4	US-09-586-472-2 Sequence 2, Appl
12	238.5	4.5	1706	4	US-09-528-706-2 Sequence 2, Appl
13	225.5	4.3	510	4	US-08-246-489-2 Sequence 2, Appl
14	225.5	4.3	2972	4	US-09-579-181-2 Sequence 2, Appl
15	225.5	4.3	3118	4	US-09-579-181-1 Sequence 1, Appl
16	224.5	4.3	533	1	US-08-040-548-1 Sequence 1, Appl
17	224.5	4.3	533	1	US-08-466-344-1 Sequence 1, Appl
18	224.5	4.3	533	6	5206152-2 Patent No. 5,206152
19	222.5	4.2	496	2	US-08-224-482-2 Sequence 2, Appl
20	222	4.2	1719	2	US-08-459-568-4 Sequence 4, Appl
21	222	4.2	1719	3	US-08-399-411-4 Sequence 4, Appl
22	222	4.2	1719	3	US-08-516-859A-4 Sequence 4, Appl
23	222	4.2	1719	4	US-09-586-472-4 Sequence 4, Appl
24	222	4.2	1719	4	US-09-528-706-4 Sequence 4, Appl
25	219.5	4.2	648	4	US-09-262-773-2 Sequence 2, Appl
26	218.5	4.1	640	4	US-09-262-773-4 Sequence 4, Appl
27	215.5	4.1	449	1	US-08-102-942A-4 Sequence 4, Appl

28	215.5	4.1	449	4	US-09-037-179B-4 Sequence 4, Appl
29	215	4.1	1234	2	US-08-317-310A-15 Sequence 15, Appl
30	215	4.1	1234	5	PCT-US95-13041-15 Sequence 15, Appl
31	214.5	4.1	429	1	US-08-234-783-4 Sequence 4, Appl
32	214.5	4.1	429	5	US-08-456-907-4 Sequence 4, Appl
33	214.5	4.1	429	5	PCT-US95-05523-4 Sequence 4, Appl
34	211	4.0	449	4	US-08-102-942A-6 Sequence 6, Appl
35	211	4.0	449	4	US-09-037-179B-5 Sequence 6, Appl
36	210.5	4.0	516	4	US-08-711-417C-202 Sequence 202, App
37	209.5	4.0	432	4	US-08-711-417C-197 Sequence 197, App
38	209.5	4.0	1581	4	US-09-110-517-2 Sequence 2, Appl
39	206.5	3.9	461	4	US-08-711-417C-196 Sequence 196, App
40	206.5	3.9	2035	1	US-08-046-585-5 Sequence 5, Appl
41	206.5	3.9	2035	5	PCT-US93-11721-5 Sequence 5, Appl
42	206.5	3.9	2035	5	PCT-US93-08743-5 Sequence 5, Appl
43	205.5	3.9	568	5	US-08-894-997-50 Sequence 50, Appl
44	205.5	3.9	976	4	US-07-945-283-2 Sequence 2, Appl
45	205.5	3.9	1958	1	US-09-449-285A-2 Sequence 2, Appl
46	204	3.9	944	4	US-08-340-203A-3 Sequence 3, Appl
47	203.5	3.9	547	1	US-08-452-567-3 Sequence 3, Appl
48	203.5	3.9	547	2	US-08-452-427-3 Sequence 3, Appl
49	203.5	3.9	547	3	US-09-085-407-3 Sequence 3, Appl
50	203.5	3.9	1291	4	US-09-150-460B-10 Sequence 10, Appl
51	203.5	3.9	1291	4	US-09-220-641-5 Sequence 5, Appl
52	203.5	3.9	1291	4	US-09-220-641-5 Sequence 5, Appl
53	203	3.8	543	2	US-08-124-482-4 Sequence 4, Appl
54	201.5	3.8	671	3	US-09-121-321-16 Sequence 16, Appl
55	201.5	3.8	671	4	US-08-933-803A-16 Sequence 16, Appl
56	198.5	3.8	711	2	US-08-820-170A-10 Sequence 10, Appl
57	198.5	3.8	711	4	US-09-055-699-10 Sequence 10, Appl
58	198.5	3.8	711	4	US-09-273-565-10 Sequence 10, Appl
59	198.5	3.8	711	4	US-09-565-538-10 Sequence 10, Appl
60	198.5	3.8	711	4	US-09-661-468-10 Sequence 10, Appl
61	198.5	3.8	1135	2	US-08-574-959A-7 Sequence 7, Appl
62	198.5	3.8	1135	2	US-09-357-014-7 Sequence 7, Appl
63	196.5	3.7	3969	4	US-08-061-376-5 Sequence 5, Appl
64	196.5	3.7	488	2	US-08-933-750C-17 Sequence 17, Appl
65	196	3.7	488	4	US-09-234-613-17 Sequence 17, Appl
66	195.5	3.7	905	2	US-08-574-959A-9 Sequence 9, Appl
67	195.5	3.7	905	4	US-09-357-014-9 Sequence 9, Appl
68	194.5	3.7	288	4	US-09-492-985-2 Sequence 2, Appl
69	192.5	3.6	518	4	US-08-711-417C-198 Sequence 198, App
70	192.5	3.6	1290	4	US-09-150-460B-6 Sequence 6, Appl
71	191	3.6	1321	2	US-08-317-310A-64 Sequence 64, Appl
72	188	3.6	675	1	US-08-317-522A-9 Sequence 9, Appl
73	188	3.6	675	1	US-08-439-818A-9 Sequence 9, Appl
74	188	3.6	675	2	US-08-751-965-9 Sequence 9, Appl
75	188	3.6	675	2	US-08-738-975-9 Sequence 9, Appl
76	188	3.6	675	2	US-08-728-626-9 Sequence 9, Appl
77	188	3.6	675	3	US-08-808-599A-9 Sequence 8, Appl
78	187	3.5	3118	2	US-08-457-273B-8 Sequence 8, Appl
79	186	3.5	1093	3	US-08-545-860D-55 Sequence 55, Appl
80	186	3.5	1093	3	PCT-US94-04496-55 Sequence 55, Appl
81	184	3.5	706	1	US-08-074-967-2 Sequence 2, Appl
82	184	3.5	706	2	US-08-553-541B-2 Sequence 2, Appl
83	184	3.5	706	4	US-09-268-202-2 Sequence 2, Appl
84	184	3.5	706	4	PCT-US94-06669-2 Sequence 2, Appl
85	184	3.5	470	5	US-08-946-241B-2 Sequence 2, Appl
86	182	3.4	470	3	US-09-309-053-2 Sequence 2, Appl
87	182	3.4	479	2	US-08-946-241B-9 Sequence 9, Appl
88	182	3.4	479	2	US-09-309-053-9 Sequence 9, Appl
89	181	3.4	376	4	US-08-874-569B-20 Sequence 20, Appl
90	181	3.4	470	2	US-08-465-590-153 Sequence 153, App
91	181	3.4	470	4	US-08-283-300A-13 Sequence 13, Appl
92	181	3.4	470	4	US-08-711-417C-153 Sequence 153, App
93	181	3.4	470	5	PCT-US95-08345-13 Sequence 13, Appl
94	175.5	3.3	1078	4	US-08-963-825-11 Sequence 21, Appl
95	175.5	3.3	1078	4	US-09-500-811-21 Sequence 21, Appl
96	175.5	3.3	1078	4	US-09-570-573-21 Sequence 21, Appl
97	175.5	3.3	1078	4	US-09-548-608-21 Sequence 21, Appl
98	175.5	3.3	2308	1	US-08-015-973-1 Sequence 1, Appl
99	175.5	3.3	2308	1	US-08-448-164-1 Sequence 1, Appl
100	175.5	3.3	2308	2	US-08-081-929-2 Sequence 2, Appl

Db 434 -----MRLETHDTK 444
 QY 523 ENTPGSEGAISGVAESSTATLTMOLSKLMTSLPSWALLTNHFKSTGSPPL----- 573
 Db 445 EHKCPHCD-----KRFNOGNLKAHLKIHADG 472
 QY 574 PLCARLALGASPSSESKLQOLVEKIDROGAVAVTSAASGAPTSAPAPSSSASSGPNOCVI 633
 Db 473 PLKCRGCGKQFTTSGNLKRLR-----IHSGEKPYVCIH 506
 QY 634 CLRVLSCPRALRLHYGQHGERPCKVCGRAFSRGNLRAHEVGH----- 679
 Db 507 CQRQFADPGLQHRHVRHTGKPCQCVWCCKAFQTQASSLIHVRQHTGKERYCERCKR 566
 QY 680 --KASPAAR-----AONSCPTCQKFTNAVTLOQHYRMLHGGQIPNGCTALPEGCGA 729
 Db 567 FVQSSOLANHIRHNDNIRPHKCSVCSKAFVNVGDSLKHITIIHG-----EKPYL 615
 QY 730 AONGSEOSTVSGAGSPFOQSOQSPPEELSEEEDEDEDEDTDEDLSLGRGSESG 789
 Db 616 CDKCGKGFNVNDLRSHVKTYHOGKAGIKIL-----EPREGSEVSVYTVDDMYT----- 664
 QY 790 GEKAISVRGDSSEASGAEVEYGTVAATAAGKEMDSNEKTTQOSSLPPPPPSLDQPOP 849
 Db 665 -----LATEALATA-----VTQLTVVP----- 682
 QY 850 MEGSSGVLGKEGKRGKPERSSSPASALTPEGEATSVTLVEELSLDEAMKEPESSSRK 909
 Db 683 -----VGAAYT--ADETEVLKAEISKAVQOVEDBNTHILY 717
 QY 910 ACEVGGAPPSQALAEHOKTH 931
 Db 718 ACDSGCGKFLDANSLAGHVRH 739
 RESULT 6
 US-08-475-844-5
 : Sequence 5, Application US/08475844
 : Patent No. 5972643
 : GENERAL INFORMATION:
 : APPLICANT: Lobanenko, Victor V.
 : APPLICANT: Neiman, Paul E.
 : APPLICANT: Klenova, Elena M.
 : APPLICANT: Goodwin, Graham H.
 : APPLICANT: Filippova, Galina N.
 : APPLICANT: Collins, Steven J.
 : TITLE OF INVENTION: CTCF
 : NUMBER OF SEQUENCES: 21
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: Townsend and Townsend Kourile and Crew
 : STREET: One Market Plaza
 : CITY: San Francisco
 : STATE: CA
 : COUNTRY: USA
 : ZIP: 94105
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: Floppy disk
 : COMPUTER: IBM PC compatible
 : OPERATING SYSTEM: PC-DOS/MS-DOS
 : SOFTWARE: Patentin Release #1.0, Version #1.25
 : CURRENT APPLICATION DATA:
 : APPLICATION NUMBER: US/08/475,844
 : FILING DATE: 07-JUN-1995
 : CLASSIFICATION: 536
 : PRIOR APPLICATION DATA:
 : APPLICATION NUMBER: 08/261,680
 : FILING DATE: 17-JUN-1994
 : CLASSIFICATION: 536
 : ATTORNEY/AGENT INFORMATION:
 : NAME: Parmelee, Steven W.
 : REGISTRATION NUMBER: 31,990
 : REFERENCE/DOCKET NUMBER: 14538A-11-1
 : TELECOMMUNICATION INFORMATION:

TELEPHONE: 206-467-9600
 TELEFAX: 415-543-5043
 : INFORMATION FOR SEQ ID NO: 5:
 : SEQUENCE CHARACTERISTICS:
 : LENGTH: 728 amino acids
 : TYPE: amino acid
 : TOPOLOGY: linear
 : MOLECULE TYPE: protein
 : US-08-475-844-5
 Query Match 4.6%, Score 241; DB 2; Length 728;
 Best local similarity 20.8%; Pred. No. 3.8e-08;
 Matches 112; Conservative 60; Mismatches 235; Indels 132; Gaps 16;
 QY 365 EKPEGRIKRCRCAYFGSDSLAQIHLRSHTGERPYKCNVCGNFTTSGNLKVHFRHREK 424
 Db 291 ERP--HKCHLCGRAFRVTLRLNHLNHTGTRPHKCPDMDMAFVTSGLY----RHR-R 342
 QY 425 YPHVQMPNRPVRENHLYVITSSGLPYGMSVPPEKAEENATPGGVBKRPVASTTALSA 484
 Db 343 YKHTHEKRFKC-SMCDYA-----SVEYSKLKRHTSHHG---ERPRQSLCSYAS 388
 QY 485 TESTLT--LSTAGTATAPGLPAENKFLYLMKAVEKRNKADENTPQSEGAISGVAESS 541
 Db 389 RDTYKLRHMRTHSGEKPEYECYICHARF-----TQSG 420
 QY 542 TATLMQSLKMTSLPSWALLTNHFKSTGSPRLPLCARALGASPSSETSKLOOLVEKIDROG 601
 Db 421 TKNHILQK-----HTENVAKFHCHCTVIRKSLGVLNLRKROHSTIEGK 467
 QY 602 AAVATSAASGAPTSAPAPSSASSGPNOCVIGLRVLSCPRALRLHYGQHGERPCKKV 661
 Db 468 KCRYCDAYFHERYALIQHQSXKNEKRFKQDCDDYACROERHVMHMRTHGKERYACSH 527
 QY 662 CGRAFSTRGNLRAHFVGHKASPAARAQNSCPTCQKFTNAVTLOQHYRMLHGGQIPNGT 721
 Db 528 CDKTFRQKOLDMHFRKRYDHPFVPAFVCSKCGKFTTRNTMARHA-----DNCS 578
 QY 722 ALPEGGAOENGSEOSTVSGAGSPFOQSOQSPPEELSEEEDEDEDEDTDEDLSL 781
 Db 579 GLDGGEG--ENGGETKKKRGKRRKRSKKESSDSDEENAEPLDONDDEE----- 627
 QY 782 AGRSESGEKAISVRGDSSEASGAEVEYGTVAATAAGKEMDSNEKTTQOSSLPPPPP 841
 Db 628 -----ETAVEIE-----AEPEVSAEAPAPPSK-----KRGGRPKGA 661
 QY 842 DSLDQPOP-----MEGSSGVLG-----KEGKPERSSSPASALTPE 880
 Db 662 TOTKOSPAALIQVEDONNGELENIVYKKEPDATVEEEDAEQAPAVEAVPNDGLTPE 720
 RESULT 7
 PCT-US95-08429-5
 : Sequence 5, Application PC/TUS9508429
 : GENERAL INFORMATION:
 : APPLICANT:
 : TITLE OF INVENTION: CTCF
 : NUMBER OF SEQUENCES: 21
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: Floppy disk
 : COMPUTER: IBM PC compatible
 : OPERATING SYSTEM: PC-DOS/MS-DOS
 : SOFTWARE: Patentin Release #1.0, Version #1.25
 : CURRENT APPLICATION DATA:
 : APPLICATION NUMBER: PCT/US95/08429
 : FILING DATE: 15-JUN-1995
 : CLASSIFICATION:
 : PRIOR APPLICATION DATA:
 : APPLICATION NUMBER: US 08/261,680
 : FILING DATE: 17-JUN-1994
 : CLASSIFICATION:
 : ATTORNEY/AGENT INFORMATION:
 : NAME: Parmelee, Steven W.

REGISTRATION NUMBER: 31,990
REFERENCE/DOCKET NUMBER: 14538A-11-1PC
TELECOMMUNICATION INFORMATION:
TELEPHONE: 206-467-9600
TELEFAX: 415-543-5043
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 728 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US95-08429-5

Query Match 4.6%; Score 241; DB 5; Length 728;
Best Local Similarity 20.8%; Pred. No. 3.8e-08;
Matches 112; Conservative 60; Mismatches 235; Indels 132; Gaps 16;

QY 365 EKGGRKRCPCAKVFSGDSALQHLRSHTEGRRPYKCNVCGNFRTRGNLKVHRRERK 424
DB 291 ERP--HKHLCGRAFRTVTLRMLNTHTRPHKCPDCDMAFVTSGLV---RRR-R 342
QY 425 YPHVMNPHVPEHLDVITSSGLPYGMSVPEPEKAEAEATPGGVERKPLVASTALSA 484
DB 343 YKTHHEKRFEC-SMCDYA-----SVEYSKLKRLHRSHTG---ERPQCSLCSTAS 388
QY 485 TESLT---LSTAGTATAPGLPAFNKFLMKAVBPKNKADENTPPGSESAISGVAES 541
DB 389 RDTYKLRHMRTHSGEKFPEYECYICHRP-----TQSG 420
QY 542 TATIMOLSKMTSLPSPNALTNHFKSTSPFLPCARALGASPSFTSKLOQLYEKIDRQG 601
DB 421 TKMHILLQK-----HTENVAKFHCPHCDYVTAARKSDLGVHRLKRSYIDQK 467
QY 602 AAVATSAASGAPTYSAPAPSSASSGPNQCVICLRYLSCPRALRLHYGOHGERPEKCV 661
DB 468 KCRVCDAVFERALALIOHSHKNEKRRKQCDYACRQERHMVMKRTHTGEKPYACSH 527
QY 662 CGRASTKGNLRHFNHGKASPAARAONSCPTQCKFTNAVTLOQHNMHLGGQIPNGGT 721
DB 528 CKTFKOKOLDHNFKRYHPNFPVPAAFVCSKCKTFTTRNRMARHA-----DNCS 578
QY 722 ALPEGGAOENSGEOSTVSGAGSFPOQSOQSPPEELSEEEEDDEEEEDVTDESL 781
DB 579 GLDGGG---ENGGETKKGRKRRKMKSKEDSDSENAEPDLDDNDEE----- 627
QY 781 AGRGSESGEKAISVRDSEASGAEEVGTVAATAAGKENDSNEKTTQOSSLPPPPP 841
DB 628 -----ETAVEIE-----AEPEVSAAPAPPPSK-----KRRGRPGKAA 661
QY 842 DSLDQPP-----MEGSSGVLG-----KEGGKPRSSSPASALTPE 880
DB 662 TOTKOSQPAIIQVEDONTGEIENIIVEVKKPEPAETVEEEEAQPAVVEAPNGDLTPE 720

RESULT 8
US-08-459-568-2
Sequence 2, Application US/08459568
Patent No. 5811304

GENERAL INFORMATION:
APPLICANT: Huang, Shi
TITLE OF INVENTION: Retinoblastoma Protein - Interacting
TITLE OF INVENTION: Zinc Finger Proteins
NUMBER OF SEQUENCES: 93
CORRESPONDENCE ADDRESS:
ADDRESSEE: Campbell and Flores
STREET: 4370 La Jolla Village Drive, Suite 700
CITY: San Diego
STATE: California
COUNTRY: USA
ZIP: 92122
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/459,568
FILING DATE: 02-JUN-1995
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/399,411
FILING DATE: 06-MAR-1995
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-LJ 1264
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1706 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-459-568-2

Query Match 4.5%; Score 238.5; DB 2; Length 1706;
Best Local Similarity 20.7%; Pred. No. 1.7e-07;
Matches 209; Conservative 97; Mismatches 319; Indels 387; Gaps 45;

QY 4 ESERSRLGVP-----AGEPAELGDDASEEDHPQVCARKCAQFTDPTTEFLAHQMAC 54
DB 828 KSEGTGTPVPWESVLDLTVAKHKPCDSEKKEKENH---LAQPAKKKKRTTCLQK--- 881
QY 55 STDPVAVVITGGQENPNNSASSSEPRREGHNPNQVMDTEHSN-----PDGSSVPTDPT 109
DB 882 -----VLL---NEYGVSLPTETTPYTRKSPCKSPDTPDPDLGPDSSCSVPTAES 931
QY 110 WGPERRGEESGHLVATGTAAGGGGLILASPKLGATPLPESTPAPPPPPP----- 164
DB 932 -PPEVVGPSPP-----LQTAISLSQLPPLLPTESSPPPCPV 972
QY 165 -----PPPGVSGHLNPLILEELRVLOQROIHQOMTEQICQVLLLSGQTVGAPAS 220
DB 973 TVATPPPLPLTPVLSHP-----SSDAS 995
QY 221 PSELPP---GTGTASTPPL-----PLFSPKPYQTSKTLAS-----SSSSSSSSSGAE 266
DB 996 PQGCPSPFSMTAQSPRLPILSPVSPSPPLPPEPLMSAASBPPTLSSSSSSSSS--- 1052
QY 267 TPKQAFPHLYHPLGSHQPFSGAGVGRSHKPTPAPSPALP---GSTDQTLAS-PLAFPT 322
DB 1053 -----FPSSSCS---STSPSPPLSAVSSVSGDNLKASLPVATFKQ- 1092
QY 323 TGLLAOCIGAAGLENTASPLGLKPKNGSGELSTGEVWGPLKPP---GGRHK----- 372
DB 1093 -----EESSEEG-LKPKR-----EAPRAGOSVQETFSKN 1122
QY 373 --CRFCAKVFGSDSALOIHLSHTGERPYKCNVCGNFRTRGNLKVH-PRH----- 421
DB 1123 FICNVCESPFLSIKDLTKHLSVNAHEMPFKCEFCVQLKVKITDLSHRRLHGVGINFVC 1182
QY 422 ---REKYPHV-QMNPHPVPEHLDVIT-----SSGLPYGMSVPEPEKA 459
DB 1183 SVCKKEFAFLCNQOQRODLHPDEVCTHHEESGTLRPQNFTDPKANEHNPSPLEPL 1242
QY 460 E-----EAPTP---GGGV-----RKPLVAS 478
DB 1243 ETSREBELNDSSDELTYTIKIMASGIKTKDPDVRGLNQHYPSFKPPFQYHHRNPMGIG 1302
QY 479 TVALS-ATESLTLSTAGATAPG-----LPANFKFVL-----MAVEEKNK---AD 522
DB 1303 VTATNFTTNIPQTFITAIKCTCGKGVNDMPLEHKLHACASASDKRYTTRKKNPVPIK 1362
QY 523 ENTTPGSESAISGVAESSTATIMQLSKMTSLPSPWALLTNHFKSTGSPFLPCARALGA 582

Db 1363 QTVQPRNGVAVLDNSGKNAFRRMGOKPRLSEFVNLGKMSPNKTL----- 1407
QY 583 SPSETSKLQOLVEKIDROGAVAVTSAAGAPTTSAAPSSASSGPNOCVLCR----- 636
Db 1408 --SALKKKNNQIVQK-----AIIQKNRAKQKADLRDTSEASS--HICPYCDREFTYIG 1456
QY 637 -----VLSGP-----RALRLHYGQHGGER----- 655
Db 1457 SLNKHAAFCSPKPKPLSPSKRVSHSKKGGHASSSSSDRNSSCHPRRRRTADTEIKMSTQ 1516
QY 656 -----PFKCKVCGRAFTSGNLR-AHFVGHKASPARAONSCPICQKKFTN-- 700
Db 1517 APLGKTRARSTGPAQASLPSSSFRSRQNVKFAASVSKKASSSLRNSSPIRMAKITHVE 1576
QY 701 -----AVTLQO-----HYRMHLGGQIPNGGTALPEGGGAQENGSEOSTV--- 740
Db 1577 GKPKRAVAKSHSAQLSSKSGRLHVYOKSKAVIOSKTAL-----ASKRRTRDFIVKSR 1630
QY 741 --SGAGSFPQOQSOQPSPEELSEEEDEDEDEE--DYTDEDSLAGSGSES 788
Db 1631 ERSGG--PITRSLQLAAADLSESRREDSARHLEKDFSYSLRLASRCGSS 1679

RESULT 9
US-08-399-411-2
Sequence 2, Application US/08399411
Patent No. 5831008

GENERAL INFORMATION:
APPLICANT: Huang, Shi
TITLE OF INVENTION: Retinoblastoma Protein - Interacting
TITLE OF INVENTION: Zinc Finger Proteins
NUMBER OF SEQUENCES: 93
CORRESPONDENCE ADDRESS:
ADDRESSEE: Campbell and Flores
STREET: 4370 La Jolla Village Drive, Suite 700
CITY: San Diego
STATE: California
COUNTRY: USA
ZIP: 92122
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/399,411
FILING DATE: 06-MAR-1995
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-LJ 1264
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1706 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-399-411-2

Query Match 4.5%; Score 238.5; DB 2; Length 1706;
Best Local Similarity 20.7%; Pred. No. 1.7e-07;
Matches 209; Conservative 97; Mismatches 319; Indels 387; Gaps 45;

QY 4 ESERSSRLGV-----AGPAPLIGDASEEDHPQVCAKCAQFTDPTPEFLAHQNAK 54
Db 828 KSEGKGRKPPVPSVLDLSVHKKPCDSEGEKFEKNH---LAQPAKAKKKPTTCMLQK--- 881
QY 55 STDPVAVIIGQENPNSSASSEPRPEGHNNPQVMDTEHSNP-----PDGSSVPTDPT 109

Db 882 -----VLL---NEYNGVSLPTETTEVTRSPSPCKSDPTQDPDELGPDSGCVPTAES 931
QY 110 WGPERRREESGHLLVAATGTAGAGGGGILLASPKLGATPLPESPAPPPPPPP----- 164
Db 932 -PPVVPSPSP-----LQTAASSGQLPPLPTPESSPPCPVYL 972
QY 165 -----PPPPVSGHNLPILEELRVLQRIHQOMTEQICRQVLLGSLGQTVGAPAS 220
Db 973 TVATPPPLPTVPLSHP-----SSDAS 995
QY 221 PSELPP---GTGTASTKPLL-----PLFSPIKPVQTSKTLAS-----SSSSSSSSGAE 266
Db 996 PQQCPSPFSNTTAAQPLPLSPVSPSPPIPEVPELMSAASGPPPLSSSSSSSSS--- 1052
QY 267 TPKAQFHLIPLPSQHPFSAGVGSRSHKPTAPSPALP---GSTDLLAS-PLHAFPT 322
Db 1053 -----FPSSSGS---STSPSPPLSAVSSVSGDNLASLPVTFKO- 1092
QY 323 TGLLAQCLGAARGLAATASPGLKPKNGSGELSYGEVMGPLKPP--GGRHK----- 372
Db 1093 -----EESSESG-LKPK-----EAPRAGGQSVVQETFSKN 1122
QY 373 --CRFCARVFGSDSALQILRSHTGERPYKCNVCGNRFTTGNLKVH-FHNN----- 421
Db 1123 FICNVCESPFLSIKDLTKHLVHAHEMPKCEFCVOLFKKYKTDLSEHNFLLHGVGNIFVC 1182
QY 422 ---REKYPHV-OMNHPRVEHLDYIT-----SSGLPYGMSVPRPKA 459
Db 1183 SVCKKEFAFLCNLOOHORDLHDEVCTHHEESGTLRPNQFTDPSKAVEMPSLPEEPL 1242
QY 460 E-----EEAATP---GGGVE-----RKPLVAS 478
Db 1243 ETSREELNDSSEELYTKIKIASGITKDPDVRGLNGHNPSPKPPFOYHNHRPMGIG 1302
QY 479 TTAALS-ATESLTLTSTAGTAPG-----LPAPNKFVL-----MKAVEPKK---AD 522
Db 1303 VTATNFTTHNIPQFTTJAIRCTKCGGVDMPELHKHILACASASDKKRYTPKKNPVPLK 1362
QY 523 ENTTPGSEGSALISGVAESSTATLMOLSKLTSLSPMALTNHFSTGSEFPLPCARALGA 582
Db 1363 QTVQPRNGVAVLDNSGKNAFRRMGOKPRLSEFVNLGKMSPNKTL----- 1407
QY 583 SPSETSKLQOLVEKIDROGAVAVTSAAGAPTTSAAPSSASSGPNOCVLCR----- 636
Db 1408 --SALKKKNNQIVQK-----AIIQKNRAKQKADLRDTSEASS--HICPYCDREFTYIG 1456
QY 637 -----VLSGP-----RALRLHYGQHGGER----- 655
Db 1457 SLNKHAAFCSPKPKPLSPSKRVSHSKKGGHASSSSSDRNSSCHPRRRRTADTEIKMSTQ 1516
QY 656 -----PFKCKVCGRAFTSGNLR-AHFVGHKASPARAONSCPICQKKFTN-- 700
Db 1517 APLGKTRARSTGPAQASLPSSSFRSRQNVKFAASVSKKASSSLRNSSPIRMAKITHVE 1576
QY 701 -----AVTLQO-----HYRMHLGGQIPNGGTALPEGGGAQENGSEOSTV--- 740
Db 1577 GKPKRAVAKSHSAQLSSKSGRLHVYOKSKAVIOSKTAL-----ASKRRTRDFIVKSR 1630
QY 741 --SGAGSFPQOQSOQPSPEELSEEEDEDEDEE--DYTDEDSLAGSGSES 788
Db 1631 ERSGG--PITRSLQLAAADLSESRREDSARHLEKDFSYSLRLASRCGSS 1679

RESULT 10
US-08-516-859A-2
Sequence 2, Application US/08516859A
Patent No. 6069231
GENERAL INFORMATION:
APPLICANT: Huang, Shi
TITLE OF INVENTION: Retinoblastoma Protein - Interacting
NUMBER OF SEQUENCES: 106

1 CORRESPONDENCE ADDRESS:
2 ADDRESSEE: Campbell & Flores LLP
3 STREET: 4370 La Jolla Village Drive, Suite 700
4 CITY: San Diego
5 STATE: California
6 COUNTRY: USA
7 ZIP: 92122
8
9 COMPUTER READABLE FORM:
10 MEDIUM TYPE: Floppy disk
11 COMPUTER: IBM PC compatible
12 OPERATING SYSTEM: PC-DOS/MS-DOS
13 SOFTWARE: Patentin Release #1.0, Version #1.25
14
15 CURRENT APPLICATION DATA:
16 APPLICATION NUMBER: US/08/516,859A
17 FILING DATE: 18-AUG-1995
18 CLASSIFICATION: 514
19
20 PRIOR APPLICATION DATA:
21 APPLICATION NUMBER: US 08/399,411
22 FILING DATE: 06-MAR-1995
23
24 APPLICATION DATA:
25 APPLICATION NUMBER: US 08/292,683
26 FILING DATE: 18-AUG-1994
27
28 ATTORNEY/AGENT INFORMATION:
29 NAME: Campbell, Cathryn A.
30 REGISTRATION NUMBER: 31,815
31 REFERENCE/DOCKET NUMBER: P-LJ 1776
32 TELECOMMUNICATION INFORMATION:
33 TELEPHONE: (619) 535-9001
34 TELEFAX: (619) 535-8949
35
36 INFORMATION FOR SEQ ID NO: 2:
37 SEQUENCE CHARACTERISTICS:
38 LENGTH: 1706 amino acids
39 TYPE: amino acid
40 TOPOLOGY: linear
41
42 MOLECULE TYPE: protein
43
44 US-08-516-859A-2

Query Match 4.5%; Score 238.5; DB 3; Length 1706;
Best Local Similarity 20.7%; Pred. No. 1.7e-07;
Matches 209; Conservative 97; Mismatches 319; Indels 387; Gaps 45;

4 ESESSRLGVP-----AGEPAELGDSSEEDHPVCAACCAQFPTPTTELAHONAC 54
828 KEEGKGTTPWMSVLDSVHKKPCDSEGEKEKNH---LAOPAAKKKKPTTCLMOK--- 881
55 STDPVAVIIGCOENPNSSASSEPRRGHNNPQVMDTEHNSP-----PDSGSSVPTDPT 109
882 -----VLL--NEYNGVSLPTTETTEVTTSPPCKSPDPTOPDELGPDSSCSVPTAES 931
110 WQPERGGESSGFLVAATGTAAAGGGGLIASPKLGATPLPEESTPAPPPPP----- 164
932 -PEVVVGPSPP-----LQTASLSSGQLPRLPTPESSPPCPVVL 972
165 -----PPPGVSGHNLPILELRLVLOQRQIHQMOMTEQICROYULLGSLQTVAPAS 220
973 TVATPTPLPLPTPLPSHP-----SSDAS 995
221 PSELP--GTGTAASSTRPL-----PLFSPKPYQTSKTLAS-----SSSSSSSSGAE 266
996 PQQCSPPFNTAOSPLPIPSPIVSPSPPIPIVPEPLMSAASPGPTLSSSSSSSSSSSS 1052
267 TPKQAFHLIHLPGSOPHSAGGVGRSHKPTPADSPALP-----GSTDQLIAS-PhLAFPST 322
1053 -----FPSSSCS--STSPSPPLSAVSSVSSGDNLASIPAVTFKQ- 1092
323 TGLLAOCIGAARGLEPAASPGLLKPKNGSGELSYGEVMGLEXP--GGRHK----- 372
1093 -----EESSEGLKPKNE-----EAPPAQGSVVOETFSKN 1122
373 --CRCRAVFGSDSALOIHILRSHGTGERPKYKNCVGNRFTTRGNLKVH-FHRH----- 421
1123 FICVNCSEPFSLIKDLTKHLVSAAEEMPFKCEFCVQLFKVATDLSSEHRLHLHGVCNIFVC 1182

QY 422 ---REKYPHV-QMNPHPPEHIDYIT-----SSGLPYGMSVPEPKA 459
DB 1183 SVCKKEFAFLNLOQHODLHDEVCETHHEFESGLRPONFTDPSKANVEHMPLEPEPL 1242
QY 460 E-----EEAATP-----GGGVE-----RKPLVAS 478
DB 1243 ETSREELNDSSEFLYTIKIMASGIKTKDPPVRLGLNQHYPSPKPPFPFYHHNNPMGIG 1302
QY 479 TPLAS-ATESITLLSTASGTATAPG-----LPAENKFLV-----MKAVEPKNK---AD 522
DB 1303 VTAVNFTTHNIPQETFTTAIRCTCKGKGVDMPELHKHILACASADKKRYTPKKNPVPLK 1362
QY 523 ENPPGSEGAISGVAESSTATIMQSLKMTSLPSSALLTHNHRKSGSPFLPLCARALGA 582
DB 1363 QTVQPKNGVYVLDNSGKAFFRMGQPKRISFVNEIGKMSPNKIKL----- 1407
QY 583 SPSETSKLOQLVERKTIDROGAVAVTSAAGCAPTTSAPAPASSASGPNOCYICLR----- 636
DB 1408 --SALKKKKNOLVCK-----AILQKNRAKOKADLRDTSEASS--HICPYCDREFTYIG 1456
QY 637 -----VLSCP-----RALRLHYGHGGER----- 655
DB 1457 SLUNKHAFSCPKKPLSPSKRKVYSHSKKGHASSSSSDNNSSCHPRRTADTEIKMSTQ 1516
QY 656 -----PRCKCKYCGNAFSTRGNLR-AHFYGHKASPARAONSCPIQCKKFTN-- 700
DB 1517 APLGKTRARSTGPAQASILPSSSFRRKRVNPKAASVSKKRASSSLNNSSPIRMAKTHIVE 1576
QY 701 -----AVTLQO-----HYRMHLGQIIPNGTALPEGGGAAGNENGSESTV--- 740
DB 1577 GKRPKAAVAKSHSAQLSSKSGRLHVQVOKSKAVIQSKTAL-----ASKRRTDRFIYKSR 1630
QY 741 --SGAGSFPOQGSQSPSEELSEEEDEDEEE--DVTDEDSLARGSEES 788
DB 1631 ERSGG--PITRSLQLAAADLSBSRRDSSARHKLDRSYSLRLASRCGSS 1679

RESULT 11
US-09-586-472-2
; Sequence 2, Application US/09586472
; Patent No. 632335
; GENERAL INFORMATION:
; APPLICANT: Huang, Shi
; TITLE OF INVENTION: Retinoblastoma Protein - Interacting
; ; Zinc Finger Proteins
; ;
; NUMBER OF SEQUENCES: 106
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell & Flores LLP
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/586,472
; FILING DATE: 01-Jun-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 09/528,706
; FILING DATE: 17-MAR-2000
; APPLICATION NUMBER: US 08/516,859
; FILING DATE: 18-AUG-1995
; APPLICATION NUMBER: US 08/399,411
; FILING DATE: 06-MAR-1995
; APPLICATION NUMBER: US 08/292,683
; FILING DATE: 18-AUG-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.

Db 932 -PPEVVGSSSP-----LQTASLSSGQLPPLTPTESSPPCPVL 972
QY 165 -----PPPGVSGSHNLPILEELRVLOQRQHOMQMTQICQVLLGLGCTVGAPAS 220
Db 973 TWAATPPPLPPLTPVLSHP-----SSDAS 995
QY 221 PSELPL---GTGTASSTRPL-----PLFSPKPYQTSKTLAS-----SSSSSSSSSGAE 266
Db 996 PQCCPSPFSMTTASQSPILSPVSPSPSPPIPVPEPLMSABSPPTLSSSSSSSSS--- 1052
QY 267 PTKQAFHLLHYPLGSHQHPFSAGVGRSHKPPRPAAPALP---GSTDQLIAS-PLHAPST 322
Db 1053 -----FPSSSCS---STSPSPPLSAVSSVSSGNDLXSLPAVTFKO- 1092
QY 322 TGLLAQCLGANGLEATASPGILKPKNGSGELSTGEVWGLEP---GGRHK----- 372
Db 1093 -----ESESSEG-LKPYE-----EAPAGQGVVQETFSKN 1122
QY 373 ---CRPQAKVFGSDSALQIHRSHTGERPKYKNCVCGNRFTTGNLKVH-FPHR----- 421
Db 1123 FICNVCESPLSLKDLTKHLSVHAEWPKCEFCYQLKVKTDLSHRPLRHGVGNIFVC 1182
QY 422 ---REKYPHV-QMNPFPVPHLDVYT-----SSGLPYGMSVPEKA 459
Db 1183 SVCKKFAFLCNDQHQRDLDHDEVCTHFEFESGTLRPQNTDPSKANVEHMSLPEPL 1242
QY 460 E-----EENATP-----GGGVE-----RKPLVAS 478
Db 1243 ETSREELNDSEELTYTTIKMASGKITKDPDVRGLGNQHPSPFPPOYHNHNPWIG 1302
QY 479 TTALS-ETSLTLSTAGTATAPG-----LPAFKFVL-----MKAVERPKN---AD 522
Db 1303 VYATNLTINIPQTFITAICTCGKGVGNMDELKHLILACASADKKYTPPKKNVPLK 1362
QY 523 EMTPEKDESSAISGVASSATITLMQLSKMTSLPSWALLTTHFKSTGSPPLICARALGA 582
Db 1363 QTVQPPGAVVLDNSGKNARRMQOPKRLSENVELGKMSPNLKL----- 1407
QY 583 SPSESKLQOLVEKIRQGAVAVTSAASGAPTTAPAPSSSSSGPNOCVICLR----- 636
Db 1408 ---SALKKKNOVLQK-----ALIQKRAAKQKADLRDTSEASS-HICPYCDREFTYIG 1456
QY 637 -----VLSCP-----RALRLHYGQHGGER----- 655
Db 1457 SLNKHAAFECPKPLSPSKRKVSHSKKGGHASSSSSDRNSCHPRRTADTEIKMQSTQ 1516
QY 656 -----PFKKVCGRAFTSTGNLR-AHFYGHKASPAARQNSCPICOKKFTN-- 700
Db 1517 APLGKTRASTGPAQASLPBSSSFRSRONVAFKFAASVSKASSSSSLRNSPDIRMAKITHVE 1576
QY 701 ---AVTLQO-----HYRMHLGQIIPNGTALPDGGAOAGSEOSTV--- 740
Db 1577 GKPKRAVAKSHSAQSSKSSRGHLHVQKSKAVYQSKTAL-----ASKRRKDRFTYKSR 1630
QY 741 ---SGAGSPFOQOQSOPSPREELSEEEDEDEEBE--DVTDEDSLARGSES 788
Db 1631 ERSGG---PIRSLQLAAADLSESRREDESSARHELKDFSYSLRLASCGSS 1679

RESULT 13

JS-08-246-489-2

Sequence 2, Application US/08246489

Patent No. 6225049

GENERAL INFORMATION:

APPLICANT: Lan, Michael S.

APPLICANT: No. 6225049kins, Abner L.

TITLE OF INVENTION: NOVEL HUMAN INSULINOMA-ASSOCIATED CDNA

NUMBER OF SEQUENCES: 27

CORRESPONDENCE ADDRESS:

ADDRESSEE: Knobb, Martens, Olson & Bear
STREET: 620 Newport Center Drive
CITY: Newport Beach

STATE: California
COUNTRY: USA
ZIP: 92660
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/246,489
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/901,715
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Israelson, Ned A.
REGISTRATION NUMBER: 29,655
REFERENCE/DOCKET NUMBER: NIH012.012A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 510 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULAR TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: N-terminal
US-08-246-489-2

Query Match 4.3%; Score 225.5; DB 4; Length 510;
Best Local Similarity 23.5%; Pred. No. 2.6e-07;
Matches 127; Conservative 46; Mismatches 232; Indels 135; Gaps 23;

QY 133 GG---GGGLIASPKLG---ATPLPEESTPAPPPPPPP-----PPPP 168
Db 22 GGEDGGRALLSSCGGARAEPAPSPVPGPLPPPPAPRAIALAALACAPGPPPPQ 81
QY 169 GVGSGHLN-----IPLLEELRVLOQRQHOM-----QMTQICQVLLGST 211
Db 82 GPRAAHFGNEAHPALVLYPTPTRVSRHEHKKYFERSFNIGSPVSAESFPPTALLGG 141
QY 212 GQTVGAPASSELPGCTASTKTLPLFSPKPYQTSKTLASSSSSSSSSSGAETPKQA 271
Db 142 G---GGGASGA---GGGTCGGDPL-LFAPAEKMGTAFSAGAEAKRGPGPPLPPAA 193
QY 272 FPHLYHPLGSHQHPFSAGVGRSHKPPRPAAPALPGSTQDLIASPHLAFSTTGLAAQCL 331
Db 194 ---ALRPPGKRRP-----PPTAAEPPA-----KAVAPAKAKKK----- 224
QY 332 GAARGL---EATASPOL-LKPKNGSELSTGEVWGLEPCKRGKRCFACAVFGSDAL 386
Db 225 -AIRKLHFEDEVTTSPVLIGLIKIKEGPVEAPRGRA-GGAARPGEPIQCLCKEYADPFL 282
QY 387 QHLRSHTGERPKYKNCVCGNRFTTGNLKVHFNHREKYYRNQMPHNPVPHLDVYTSS 446
Db 283 AQHKCSRIVREYKCECAKAVFSCPANLASHRRWHKPR-PAPAAARADPE-----AA 334
QY 447 GLPYGMSVPEKAEEBAATPGGVE-----RKPLVASTTALS 483
Db 335 ARAEAREAPGGGSDRDTSPGCVSESESDGLYECNHCAKKFRQAYLRKHLAHNQALQ 394
QY 484 ATESLTILSTAGTATAPGILPAFNKLYLMAKAVEPKKADENTPPGSESAISGVASSSTA 543
Db 395 A-----KGAPLAP--PAEDLIALVPG--PDEKAPQEAAGDGEAGVIGLSASABC 440
QY 544 TLMQLSKMTSLPSWALLTTHFK---STGSPPLPICARALGASPESTKIDQLVKIDRO 600
Db 441 HICPVCG---ESFRASKGAQERHRLHLHAQVFPCKYCPATPTVSSGTLTHINKCHPSENRQ 498


```
RESULT 14
US-09-579-181-2
; Sequence 2, Application US/09579181
; Patent No. 6365372
; GENERAL INFORMATION:
; APPLICANT: Chivita, John
; APPLICANT: Yaciuk, Peter
; TITLE OF INVENTION: SNF2 Related CBP Activator Protein (SRCAP)
; FILE REFERENCE: 16153-4247
; CURRENT APPLICATION NUMBER: US/09/579,181
; CURRENT FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: 60/136,620
; PRIOR FILING DATE: 1999-05-27
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 2
; LENGTH: 2972
; TYPE: PRT
; ORGANISM: Human
US-09-579-181-2

Query Match      4.3%; Score 225.5; DB 4; Length 2972;
Best Local Similarity 21.0%; Pred. No. 2.5e-06;
Matches 225; Conservative 101; Mismatches 366; Indels 377; Gaps 50;

OY 58 PVMVITIGQENPNNSASSEPR-----PEGHNN---POVMDTEHSNPPDSGSS 103
    | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 890 PSLGLVSGTSRPTPTLSLKPTPPAVRLSPAPRPGSSSLKPLTVPPGYTFPPAAATT 949
    | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
OY 104 VPRD-----PTWGPBR-----RGE-ESSGHFLVAATGTAAGGGG-- 137
    | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 950 TSTTTATTATTAVPAPTPAPQRLILSPDMQARLPSCGEVVSIGOLASLAQRPAVNAAGSKP 1009
    | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
OY 138 -----LILASPKIGATPLPPES 154
    | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 1010 LTFQIQGNKLLTGAOVROLAVGPRPLQMPPTMVNNTGVYKIVVRQAPRDLTPVP--L 1068
    | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
OY 155 TPAPPP-----PPP-----PPRGVSGHLNIPILIEELRVLOQRQIHQMOMTE 199
    | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 1069 APAPRPPSSGLPAVLNPRPTLTGRLPTTLGTARAMP-----TP 1109
    | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
OY 200 QICROVLL-----GSLGQTVGAPAS--PSELPGTGAASKTKPL---LPFSP 242
    | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 1110 TLVRPLKLIVHSPREVSASAPGAAPLTITSSPLHVSSLP--GPASSPMPIPNSSPLASP 1167
    | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
OY 243 IKPVQTSKTIASSSSSSSSSGAET--PKQAFHLVHPILGSHPFSGAGVGRSHKPTPAPS 301
    | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 1168 V-----SSTVSVPLSSSLPISVPTLLPAPASAPLTIPISAPLTVSASGAPALLTSVTPPLA 1222
    | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
OY 302 PALPGSTDLIASPHLAPRSTTGLLAOCLGARGLEATASPG--LLKPKKSGSELSYGE 359
    | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 1223 PVPAPAGPPSLQPGGAPSPASALTLG--LATAPSLSSSOTPGHPLLLAPTSIHVGLNS 1280
    | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
OY 360 VMGLEKPGGHHKCFKCAKVFSGDSALQIHLRSHTGERPYKCNVCNFTTRGNLKVHFH 419
    | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 1281 TVARA-----CSRVVLPASAL----- 1296
    | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
OY 420 RHREKYPHVQMNPHREHLDVYTSSGLPYGMSVPEPEKAEEEAATPGGVERKPLVAST 479
    | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 1297 --ASPFPSA---PNPAPAA-----SLIAPASSASQALATP-----LAPMAAPQ 1335
    | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
OY 480 TALSATESLTLSTAGATATAGLPANFKFVLKMKAVPEPKKADENTPBGSSGSAISGV-- 537
    | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 1336 TAILAPSPAPPLAPPLVAPSPQAAD---VLASSQTPVPVPMADSTPGTSLASASVPA 1391
    | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
OY 538 -----AESSTALMOLSKIMTSLPSWALLTNHFKSTGSFPL--PLCARAL--GASPESTKYL 590
    | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 1392 PTPVLAPBSTQIMLP--APVPSPLPSA-----STOTIALAPALAPTLGGSSPQOTLSL 1443
    | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
OY 591 QQLVEKIDRQ-----AAVATSAAGAPTTS-----APASSASSSGPNQVCYICLRVLSC 640
    | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 1444 ----GTGNQDQGFPTOTLSLTLPASSSLVPTPAQTLSLAPGP-----PLGPTQ-----TLSL 1489
    | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
```

```
OY 641 PRALRLHYGQHGEBRPPCKKVCGRAFSTRGNLRAHFVGHKASPARAONSCPICQKKFTN 700
    | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 1490 APAPPLPASPVPAP-----ARTL-----TLAPASSASLILAPASVQ 1527
    | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
OY 701 AVTLQOHVRHMLGQIIPNGTALPEGGAQDENGSEOST--VSGAGSFPOQSOQSPPEE 758
    | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 1528 TULTSPAPVPTLG---PAAQTLALAPASTQSPASQASSLVVASGA-----APLPVT 1577
    | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
OY 759 ELSEEEEEDEEEDVDYDESLAGR-GSESGEKALISYRGDSDEASGAEETVAAAA 817
    | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 1578 MVSRLPVSKDE-----PDTLTLRSGPPSPSTATSGGPPRRQ----- 1616
    | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
OY 818 TAGKEMDSNEKTTQSSLP PPPP-----DSLDPQPMQ-----GSSGV 857
    | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 1617 -----PPPPRSPFLTDLSEKRRKNORSRLERIFQLSEAHGALAP 1657
    | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
OY 858 LGGKEE---GKKPERSSPASALTP-----EGEATSVTLVEEL--SLQEMRK- 900
    | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 1658 VYGTREVLDFTLPQPVASPIGPRSPGSHPTWYTEAHARAVLFPQQRDLQLSIIEIRF 1717
    | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
OY 901 ----EPGES--SSKACEVCGQAFPSQALDEH--QKTHKEGRLFTCY 941
    | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 1718 IFVMPVEAPRPPSLHACHPPWLPARQAAFOEQLASELMPRARLRIV 1766
    | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
```

```
RESULT 15
US-09-579-181-1
; Sequence 1, Application US/09579181
; Patent No. 6365372
; GENERAL INFORMATION:
; APPLICANT: Chivita, John
; APPLICANT: Yaciuk, Peter
; TITLE OF INVENTION: SNF2 Related CBP Activator Protein (SRCAP)
; FILE REFERENCE: 16153-4247
; CURRENT APPLICATION NUMBER: US/09/579,181
; CURRENT FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: 60/136,620
; PRIOR FILING DATE: 1999-05-27
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 1
; LENGTH: 3118
; TYPE: PRT
; ORGANISM: Human
US-09-579-181-1

Query Match      4.3%; Score 225.5; DB 4; Length 3118;
Best Local Similarity 21.0%; Pred. No. 2.6e-06;
Matches 225; Conservative 101; Mismatches 366; Indels 377; Gaps 50;
```

```
OY 58 PVMVITIGQENPNNSASSEPR-----PEGHNN---POVMDTEHSNPPDSGSS 103
    | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 1036 PSLGLVSGTSRPTPTLSLKPTPPAVRLSPAPRPGSSSLKPLTVPPGYTFPPAAATT 1095
    | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
OY 104 VPTD-----PTWGPBR-----RGE-ESSGHFLVAATGTAAGGGG-- 137
    | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 1096 TSTTTATTATTAVPATTPAPQRLILSPDMQARLPSEGEVVSIGOLASLAQRPAVNAAGSKP 1155
    | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
OY 138 -----LILASPKIGATPLPPES 154
    | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 1156 LTFQIQGNKLLTGAOVROLAVGPRPLQMPPTMVNNTGVYKIVVRQAPRDLTPVP--L 1214
    | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
OY 155 TPAPPP-----PPP-----PPRGVSGHLNIPILIEELRVLOQRQIHQMOMTE 199
    | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 1215 APAPRPPSSGLPAVLNPRPTLTGRLPTTLGTARAMP-----TP 1255
    | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
OY 200 QICROVLL-----GSLGQTVGAPAS--PSELPGTGAASKTKPL---LPFSP 242
    | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 1256 TLVRPLKLIVHSPREVSASAPGAAPLTITSSPLHVSSLP--GPASSPMPIPNSSPLASP 1313
    | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
OY 243 IKPVQTSKTIASSSSSSSSSGAET--PKQAFHLVHPILGSHPFSGAGVGRSHKPTPAPS 301
    | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
```


Db 1314 V-----SSTVSVPLSSSLPISVPTLLPAPASAPLITIPISAPLTVSASGPAALLTSVPLEA 1368
QY 302 PALPGSTDOIASPHLAFSTTTGILAAOCLGARGLEATASPG--ILKPRNGSGELSTGE 359
Db 1369 PVVPAAGPSPLOSPGASPSASALTIG--LATAPSSSSQTPGHPLLATPTSSVHPGINS 1426
QY 360 VMGPLEKPGGRHKRCFCAKVFSGSDALOIHLSRHNGERPCKVNCGRNFTTRGNLKVHFH 419
Db 1427 TVAPA-----CSPVLVPAASAL----- 1442
QY 420 RHHEKYPHOMNHPVNEHLDYVITSSGLPYGMSVPPEKAEERATPPGGVEKKPLVAST 479
Db 1443 --ASPPSA--PMPAPAQA-----SLAPASSAQALAP-----LAPMAPQ 1481
QY 480 TALSTATESLTLSTASGATAPGLPAFNKFMKAVEPKNKADENTPPGSESAISCV-- 537
Db 1482 TALAPSPAPPLAPPLAPLAPSPGAAP-----VLASSTPPVMAVPSSTPGTSLASASVPA 1537
QY 538 -----AESSTATLMQSLKMTSLPSSWALLTNHFKSTGSEPL-PLCARAL-GASPSSETSKL 590
Db 1538 PVPVLAAPSSQTMLP-APVSPPLPSPA-----STQTLALAPALAPLTGSSPSQTLSL 1589
QY 591 QOLVERIDRG-----AVAVTSASGAPTS-----APAPSSASSGPNOCVLCRLVLS 640
Db 1590 -----GTGNPQGGPPTQTLSLTPASSLVPTPQOTLSLAPGP--PLGPPQ-----TLSTL 1635
QY 641 PRALRLHYGQHGGERPKCKYCGRAFSTRGNLRAHFVGHKASPARARONSCPICOKKFTN 700
Db 1636 APAPPLAPASPPVAPAP-----AHTL-----TLAPASSASLILAPASVQ 1673
QY 701 ATTLQOHVNMHLGGQIPNGSTALPEGGGAOENGSEOST--VSGAGSPQOQSOQPSPEE 758
Db 1674 TLTLSAPVPTLd--PAAAOITALAPASTQSPASQASLLVVASAGA-----APLPVT 1723
QY 759 ELSEEEDEDEEEDVTDSDSLAGR-GSESGEKAISVRGQSESEASGAEEVGYAANA 817
Db 1724 MVSRLPVXND-----PDTLTLKSGRPSPPTATISFGGRRPRKQ----- 1762
QY 818 TAGKEMDSNEKTTQOSSLPRLPPPP-----DSLDPQPMEMQ-----SSSGV 857
Db 1763 -----PPPRSPFYLDLSLEKRRKRQSERLERITQLSFAHGLAP 1803
QY 858 LGGKEE--GKPERSSSPASALTP-----EGEATSVTLVEEL--SLQEMARK- 900
Db 1804 VGTGYLDFTCLPQVAVASPIGSPSPGSHPTFWTYTEAHRVNLFPQOQLDQISEITIERP 1863
QY 901 ---BPGES--SSRKACEVGOAFPSQALAEH--OKTHPKEGPLFTCV 941
Db 1864 IFVMPVPEAPPPSLHACHPPRWLAPRQAAPQOQLASELMPRARPLHRTV 1912

RESULT 16
US-08-040-548-1
; Sequence 1, Application US/08040548
; Patent No. 5763209
; GENERAL INFORMATION:
; APPLICANT: Sukhame, Vikas P.
; TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO THE
; TITLE OF INVENTION: FUNCTIONAL DOMAINS OF DNA BINDING PROTEINS
; NUMBER OF SEQUENCES: 67
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Arnold, White & Durkee
; STREET: 321 No. 5763209th Clark Street, Suite 800
; CITY: Chicago
; STATE: Illinois
; COUNTRY: U.S.A.
; ZIP: 60610
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA: 1

APPLICATION NUMBER: US/08/040,548
; FILING DATE: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Coughlin, Daniel F.
; REGISTRATION NUMBER: 36,111
; REFERENCE/DOCKET NUMBER: arc0067
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (312) 744-0090
; TELEFAX: (312) 245-4961
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 533 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-040-548-1

Query Match 4.3%; Score 224.5; DB 1; Length 533;
Best Local Similarity 23.0%; Pred. No. 3.2e-07;
Matches 135; Conservative 60; Mismatches 190; Indels 201; Gaps 28;

QY 99 DSGSVPTDPTWGPBERGEE-----SSG--HFLVAA-----TGTAAAGGGGLILA 141
Db 18 DPGSPFPHSPMTMDYPRKLEEMMLLSNGAPQFLGAGTPBGSGGSSSTSSGGGG--G 74
QY 142 SPKLGATPLPESTPAPAPPPPPPPPPPPGVGSHLNPILLELRVLAQOROIHOMQMTQI 201
Db 75 GSNSSGAENPQGPBSPQ-----YEHLTTE-SFSDIALNNEKAMVETSYSPSQ 122
QY 202 CROVLLIGSGOVYGAAPASPELPGTGTASTKPLPLPSPIKPY--QTSKTLASSSS 258
Db 123 TR--LPRTTYGREFLEPA--ENSGTLMPEPLPSLVSGLYSKTNPTSSSASBPAA 176
QY 259 SSSSSGAETPKQAFHLYHNPGLSQHPESAGVGSRSHKPT-PAP-----SPALPGST 308
Db 177 SSSSSASQSP-----PLSCAVPSNDSSPIYSAAPTFPPRNTDIFEPDQQAAPGSA 227
QY 309 DQILASPHLAFPSITG-----LLAAQ-----CLGNA-----RGLE-ATASPLG--- 345
Db 228 GTALQYPPPAVPATKGGFOVPMIPDYLEPQOQGLSLCTPDQKPPQGLNRRNQPSLITPL 287
QY 346 ----LKPNGSGEL-----SYGEVM-----GPLEKPGG----- 369
Db 288 STTKAFATQSGQDKALNTTYQSLIKPSRMKRYPNRPSKTRPHRRPTACPYESCDRRF 347
QY 370 -----RH-----KRCFCAKVFSGSDALOIHLSRHTGERPKYKVCNGNREPTRG 412
Db 348 SRSDLLTRHTRHTGOKRPFQCRICMRNPSRSDHLTTHTRHTGKERPFACDICGRKPARSD 407
QY 413 NLKVHFHRH-REKYRPHQNMNHPVNEHLDYVITSSGLPYGMSVPPEKAEERATPPGGVE 471
Db 408 EKKRHHTKIHROK-----DKKAQ----- 425
QY 472 RPELVASTALSTATESLTLSTASGATAPGLPAFNKFMKAVEPKNKADENTPPGSE- 530
Db 426 -KSVASPAASS-----LSSYPSVATSYSPATTSF-----PSVPTSYSSPGSSST 471
QY 531 --GSAISGVAESSTALUM-----QLSKLMTSLPMSWALLTNHFKSTG 569
Db 472 YPSPAHSGFPPSPSVATTFASVPPAFTPQVSSFPAGVSSSSTSTSG 517

RESULT 17
US-08-466-344-1
; Sequence 1, Application US/08466344
; Patent No. 5773583
; GENERAL INFORMATION:
; APPLICANT: Sukhame, Vikas P.
; TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO THE
; TITLE OF INVENTION: FUNCTIONAL DOMAINS OF DNA BINDING PROTEINS
; NUMBER OF SEQUENCES: 67

GENERAL INFORMATION:
APPLICANT: Mercola, Dan
APPLICANT: Adams, Eileen D.
TITLE OF INVENTION: Inhibition of the Mitogenic Activity of
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESS: Campbell and Flores
STREET: 4370 La Jolla Village Drive, Suite 700
CITY: San Diego
STATE: California
COUNTRY: USA
ZIP: 92122
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/224,482
FILING DATE: 07-APR-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-ME 9913
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 496 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-224-482-2

Query Match 4.2%; Score 222.5; DB 2; Length 496;
Best Local Similarity 22.8%; Pred. No. 4e-07;
Matches 127; Conservative 58; Mismatches 180; Indels 191; Gaps 26;

112 PERREESSGHFLVATGTAGGGGLIASPKIGATPLPESTPAPPPPPPPPPG 171
18 PEGSGNGS-----SSSSSGGGG---GSGNSGSAFNPQGESECP----- 56
172 SGHNLPILEELRLQROHQOMTEQICROYLLGSLGQTVGAPASPSLPGTAS 231
57 YEHLTTE-SFDIALNNEKAMVETSPSQTR---LPITYTGRFSLEPA--PNSGNTL 109
232 STKPLPLESPKPV---QTSKTLASSSSSSSGAETPPKQAFPHLYHPLGSHPEFSAG 288
110 WPEPLSLVSGVSMTPPTSSSSAPSPASSSSSSASQSP-----PLSCAVPSNDS 160
289 GVGSHKPT-PAP-----SPALPGSTDLIASPHLAPPSSTG-----LLAA 328
161 SPVSAATFPPNNDIFEPGQAFPGSAGTALQYPPPAITKGGQVPPHPIPTLFPQ 220
329 Q-----CLGAA-----RGLE-ATPASPGL-----LKPKNSSGEL-----SYGEVW----- 361
221 QQGDLSLCTPDQKPGQLENRTQPSILPLSTJKAFAMQSGSODLKALNTYQSLIKPS 280
362 -----GPLEKPG-----RH-----KRFCAKVGCS 382
281 RMKRYPPNPSTPPHERPACPVESCDRRFSRDELTNRHRIHTGOKPFOCRICKMNSR 340
383 DSALQIHLSHTGEPHYKCNVGNRFTTNGNLKVHFHRH-REKYPHYQMNHPVREHLDY 441
341 SDHLTHTHTGEPFACDICKGRKARSDERKRTIHLRQ----- 383
442 VITSSGLPYGMSVPEKAEAEATPPGGVEKPKLVASTALSTATESLTLLSTASCTATAP 501
384 -----DKRAD-----KSVVASPAASS-----LSSYPVVAITSY 411

502 GLPAFNKFLMKAVEPKKADENTPGSF---GSAISGVAESSTATLM-----QLSKLMT 553
412 PSPATYSF-----PSPVPTSPSGSSTYSPSAHSGPPSPVATTFAPVPAFPVOVS 464
554 SLPSWALLTNHFKSTG 569
465 SPFSAGVSSSPSTSTG 480

RESULT 20
US-08-459-568-4
Sequence 4, Application US/08459568
Patent No. 5811304
GENERAL INFORMATION:
APPLICANT: Huang, Shi
TITLE OF INVENTION: Retinoblastoma Protein - Interacting
NUMBER OF SEQUENCES: 93
CORRESPONDENCE ADDRESS:
ADDRESS: Campbell and Flores
STREET: 4370 La Jolla Village Drive, Suite 700
CITY: San Diego
STATE: California
COUNTRY: USA
ZIP: 92122
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/459,568
FILING DATE: 02-JUN-1995
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/399,411
FILING DATE: 06-MAR-1995
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-LJ 1264
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1719 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-459-568-4

Query Match 4.2%; Score 222; DB 2; Length 1719;
Best Local Similarity 17.8%; Pred. No. 2.1e-06;
Matches 232; Conservative 142; Mismatches 425; Indels 502; Gaps 53;

9 SRUGVPAGEPAELG-----GDASEED-----HPVCAKCAQGTDPTE--F 47
40 TRIGVWATPIILKKGFGFVGDKKKRSQYKNNVYMEVYIPNL-GWMCIDATIDBEKGMW 98
48 LAHON-ACS-----TDPV-----MTIIGOENPNSSASSEPRPE 82
99 LRYVMAGSGEONLFLPILNRAIYYKTLKPIAFGEELVWYNGEDNDEIAAIEERAS 158
83 G-----HNNPQVMTPE-HSNPDSGSSVTPDPVWGPERRKEESSG 121
159 ARSKRSSPKSRGKKKSQENKKNKIKQIDQLKTSBDDFTSANNRDSEAGPREDKEKPSA 218
122 HFLVATGTAAAGGGGLIASPKIGATPLPE-STPAPPPPPPPPPPPGSGHNLNPLI 180
219 SALDEPA-----TLQEVASQEVPELATPPAPAMEPQEPDERLERAAACEVNDL 266
181 LLELRVLIQORLIHQOMTEQICROYLLGSLGQTVGAPASPSLPGTASSTKPLDLF 240

Db 267 GEE-----EEEEEEEEEE-----DDDDDELEDEGEFEASMP----- 300
QY 241 SPIKPVOTSKTIASSSSSSSSGAEPTPKQAFPHLYHPLGSHPSAGVGSRHKPTPAP 300
Db 301 -----NENSVKEPE-----IRCDKEP----- 316
QY 301 SPALPGSTQOLIASPHLAFSTTGLLAQCLGAARGLE--ATASPGLLKPKNGSGELSYG 358
Db 317 -----EDLLEPKTJSEET-----LEDCESEVTPAMQIPIR--TKEEANG 352
QY 359 EVMGPLEKPGGRHKCRFCAGVFGSDSALOIHLSRH--TGERPYKCVGNGRFTTGMILKY 416
Db 353 DVEFTTMMF-----CQCEKERTTKQGLERHMHIIHISTVNHAFKCKYCGKAFGTQIARR 407
QY 417 HFRHRH--EKYPHVQNNPHRPHLDVYVITSSGLPYGMSV-----PP----- 456
Db 408 HERRHAGLKRKPSOTLQIP--SEDL-----ADGKASGENVASKDDSSPPLGPDCLIMN 459
QY 457 -EKAEEEAATPGGVEKRPVLAATSTALATESLTLSTAGTAT--APGLPAFNKFIYM 512
Db 460 SEKASODTI-----NSSVVEENGVEKELHPCKYCKKVGTHTNMRRHORVRHERHLIP 512
QY 513 KAVEPKKADENTPGSEGSALISGV----- 537
Db 513 KGVRRKGLEPPQAPQAOATQNVYVSTEEPEEGEADVYIMDISNTISENLNYIDG 572
QY 538 -----ABSSATATLMOLSKLMTSLPSWALLTNHKS----- 567
Db 573 KIOTNNNTSNCVDYIEMESADLVGINCLLT--PYVEITQNIKTQVPTEDLPKEPLG 630
QY 568 -----TGSFPL--PICARALGASPSFT-----SK 589
Db 631 STNSEAKKRRRTASPPALPKIAETISDPPVPCGSLPLSISTTEAVSFHKEKSVYLSK 690
QY 590 LQOLVE-----KIDROGAAVNTSAAGAPTSA-----PAFSSASGPN 629
Db 691 LKQLOLQODKLTTPRAGISATEIATLGPVCVSAAPASMLPVTSRFRKRTSSPPSQHSP- 749
QY 630 QCVYCLAVLSPRALRLHYGCHGERPFKCKYCGRAFSTRGNL--RAHFVGHASPA-- 684
Db 750 -----ALR--DFGRPSD-----GKAMWDAGLTKSKSKLSDSDSPAMS 786
QY 685 -----ARAONSCPIC-----OKFT-----NAVTLQOHVHMLGQIPINGST----- 721
Db 787 LSGDERETVSPPCDEYKMKSEMTASSAFSSVCNQPLDLSSGVKKAEGTGKTPIVQWE 846
QY 722 -----ALPEGGGAOENGSEOSTVSGAG-----SF 746
Db 847 SVLDLSVHKHCHSDSEGEKFEKESHVQPTCSAVKRRKPTTCMLQKVLINEXNGIDLPEVN 906
QY 747 PQOOSQOPSPBEELSEEEEEE-----DEEBEDYDDEDSLACGSRSGE--KAIS 795
Db 907 PADGTRSPSPCKSLAQDPDLGPGSGFPAPVVESTPDVCPSSPALOTPSLSSGOLPIL 966
QY 796 VRGD-----SEASGAEEBVGTVAAATAGKENDSMEX 828
Db 967 IPTPSSPPPCPVLTVAITPPPLIPTVLPAPSSASAPHCPCPLSNATQOSPILISP 1026
QY 829 TTQOSSLPPPP--PPDSLDQPPPMEOGSGCVLGKKEGKPERSSS-----PASKLTP-- 879
Db 1027 TVSPSPSPPIPVPEPLMASAPGPPTLSSSSSSSSSSSSSSSSSPSPPLSASSIVY 1086
QY 880 -----EGEATSVTLVEELSLOEAM--RKEPGESSSRK-----ACEVCOAFPS 920
Db 1087 SSGGNLEASLPMSIKOELENGELKPREPQSAADQVVOETFNKNFVNCVSCSPFLS 1146
QY 921 QAALEEHOKTHPEKGPPLFTVCRCROGFLEERATLKKHMLAH 961
Db 1147 IKDLTKHLHSIHAEWP--FKCEFCVOLFKDKTDLSEHRLHL 1186
RESULT 21

US-08-399-411-4
; Sequence 4, Application US/08399411
; Patent No. 5831008
; GENERAL INFORMATION:
; APPLICANT: Huang, Shi
; TITLE OF INVENTION: Retinoblastoma Protein - Interacting
; TITLE OF INVENTION: Zinc Finger Proteins
; NUMBER OF SEQUENCES: 93
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell and Flores
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/399,411
; FILING DATE: 06-MAR-1995
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 1264
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1719 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-399-411-4
Query Match 4.2%; Score 222; DB 2; Length 1719;
Best Local Similarity 17.8%; Pred. No. 2,1e-06;
Matches 232; Conservative 142; Mismatches 425; Indels 502; Gaps 53;
QY 9 SRLGVPAEPALG-----GDASEED-----HPQYCAKCAOFTDTE--F 47
Db 40 TRIGVWATKPIIKGKFKGPFVGDKKRSQYKNNYMEVYYPNL--GWMCIDATDEKGNW 98
QY 48 LAHON-ACS-----TDPPV-----MYIIGOENNNSSASSEPRPE 82
Db 99 LRYVWACSGEBONLFPLEINRAIYKTKLPVAPGEBLLVWYNGEDNPEIAAIEERAS 158
QY 83 G-----HNNPQVMDTE--HSNPPDSGSSVPTDPTWGPERRGEESG 121
Db 159 ARSKSSSPKSRGKKKSKOENKNGKNIQDIQLKSEPDFTSANMMDSABGPREDEKPSA 218
QY 122 HFLVAATGTAAGGGGLIASPKLGATPLPPE--STPAPPPPPPPPPPVGSGHLNIPLI 180
Db 219 SALEQPA-----TLQEVASQEVPELATPPAPMBPOPPDRDLLEAAACEVNDL 266
QY 181 LEELAVLQORQIHOQMTEDICROVLLGSLGVGAVAPASSELPGTGTASTKPLPLPFE 240
Db 267 GEE-----EEEEEEEEEE-----DDDDDELEDEGEFEASMP----- 300
QY 241 SPIKPVOTSKTIASSSSSSSSGAEPTPKQAFPHLYHPLGSHPSAGVGSRHKPTPAP 300
Db 301 -----NENSVKEPE-----IRCDKEP----- 316
QY 301 SPALPGSTQOLIASPHLAFSTTGLLAQCLGAARGLE--ATASPGLLKPKNGSGELSYG 358
Db 317 -----EDLLEPKTJSEET-----LEDCESEVTPAMQIPIR--TKEEANG 352
QY 359 EVMGPLEKPGGRHKCRFCAGVFGSDSALOIHLSRH--TGERPYKCVGNGRFTTGMILKY 416

Db 353 DVFETMP-----CQHCERKFTTKQGLERHMHIIHSTVNHAFKCKYCGKAFQTQINRR 407
QY 417 HFHRHR---EKYPHVMNHPVPEHLDYVITSSGLPYGMSV-----PP----- 456
Db 408 HERRHEAGLKRKPSQTLP-----SEDL-----ADGKASGENVASKDSSPSLGPDLIMN 459
QY 457 -EKAEEBAATPGGVERKPLVASTTALSTATESLTLTSTAGTAT---APGLPAFNKFLVM 512
Db 460 SEKASODTI-----NSSVVEENGEVKELHPCYCKYKVFQHTNNRRHQRVHERHLIP 512
QY 513 KAVEPNKADENPPGSEGSALISCV----- 537
Db 513 KGVRRKGGLEEPPAPAOAQTONVYVPSTEEPEEGEADVYIMDISNISENLNYIDG 572
QY 538 -----AESSTATLMQSLKMTSLPSMALLTNHFKS----- 567
Db 573 KIOTNNNTSNCVYIEMESASADLGINCLLT--PYVETTONIKTQVPTEDLPKEPLG 630
QY 568 -----TGSFPL-PLCARALGASPSPT-----SK 589
Db 631 STNSEAKKRTASPPALPKIKAEFTSDPMVPSCSLSLPLSISTTEAVSFHKEKSVYLSK 690
QY 590 LQOLVLE-----KIDROGAVAVTSAASGAPTTSA-----FAPSSASGPN 629
Db 691 LKOLLOTQDKLTTPAGISATEITAKLGPVCSAPASMLPVTSSRFKRRRTSSPPSPQHP- 749
QY 630 QCVICLRVLSCPRALRLHYGOHGEGRPFCKVCGRAFSTRGNL---RAHFVGHKASPA-- 684
Db 750 -----ALR-DFGKPSD-----GKAAMTDALITSKSKLSHSDSPAWS 786
QY 685 -----ARAONSCPTC-----QKFT-----NAVTLQOHVRHMLGGOIPNGGT----- 721
Db 787 LSGRDERETVSPCEFEDEYKMSKEMTASSAFSSVCNOQPLDLISGVKQKAEGTGTPVQWE 846
QY 722 -----ALPEGGGAOENGSQSTVSAG-----SF 746
Db 847 SVLDLSVHKHHCSDSEKREKESHVSOPTCSAVYKRRKPTTCMLQVLLNEVNGIDLPEVN 906
QY 747 POOQSOQSPSEELSEEEBE-----DEBEEDVTDEDSLARGSESGE--KAIS 795
Db 907 PADGRSPSPCKSLAOPDPDLPGSGFPATVSTPVCSSPALQPTSLSSGQLPPL 966
QY 796 VRGD-----SEBASGAEEVGYVAAAATAGKEMDSNEK 828
Db 967 IPTDSSPPPCPVLTVAAPPPLLPVPLPAPSSSASPNCPSPLSNATAOSPLILSP 1026
QY 829 TTOQSSLPPPP--PPSDLOPQPMEOGSSGVLGKEGKPERSSS-----PASALTP-- 879
Db 1027 TVSPSPPIPVPEPLMSASAPGPTLSSSSSSSSSSSSSSSSSSSSSPPLSLAISVY 1086
QY 880 -----EGEATSVTLVEELSLQELAM--RKEPGESSSRK-----ACEVCGOAFPS 920
Db 1087 SSGDMLASLPMISFKOELENGELKPREPQSAEQDVVYQETFNKNFVNCVCESPFIS 1146
QY 921 OAALEEHQKTHPKEGPLFCVPCROGFLERATLKKHMLAH 961
Db 1147 IKDLTKHLISHAEWF-FKCERCVOLFKDKTDLSEHAPFLH 1186

RESULT 22
US-08-516-859A-4
Sequence 4, Application US/08516859A
Patent No. 6069231

GENERAL INFORMATION:
APPLICANT: Huang, Shi
TITLE OF INVENTION: Retinoblastoma Protein - Interacting
NUMBER OF SEQUENCES: 106
CORRESPONDENCE ADDRESS:
ADDRESSEE: Campbell & Flores LLP
STREET: 4370 La Jolla Village Drive, Suite 700
CITY: San Diego
STATE: California

/?
/? COUNTRY: USA
/? ZIP: 92122
/?
/? COMPUTER READABLE FORM:
/? MEDIUM TYPE: Floppy disk
/?
/? COMPUTER: IBM PC compatible
/? OPERATING SYSTEM: PC-DOS/MS-DOS
/? SOFTWARE: PatentIn Release #1.0, Version #1.25
/?
/? CURRENT APPLICATION DATA:
/? APPLICATION NUMBER: US/08/516,859A
/? FILING DATE: 18-AUG-1995
/? CLASSIFICATION: 514
/?
/? PRIOR APPLICATION DATA:
/? APPLICATION NUMBER: US 08/399,411
/? FILING DATE: 06-MAR-1995
/?
/? PRIOR APPLICATION DATA:
/? APPLICATION NUMBER: US 08/292,683
/? FILING DATE: 18-AUG-1994
/?
/? ATTORNEY/AGENT INFORMATION:
/? NAME: Campbell, Cathryn A.
/? REGISTRATION NUMBER: 31,815
/? REFERENCE/DOCKET NUMBER: P-LJ 1776
/?
/? TELECOMMUNICATION INFORMATION:
/? TELEPHONE: (619) 535-9001
/? TELEFAX: (619) 535-8949
/?
/? INFORMATION FOR SEQ ID NO: 4:
/? SEQUENCE CHARACTERISTICS:
/? LENGTH: 1719 amino acids
/? TYPE: amino acid
/? TOPOLOGY: linear
/?
/? MOLECULE TYPE: protein
/?
/? US-08-516-859A-4

Query Match 4.2%; Score 222; DB 3; Length 1719;
Best Local Similarity 17.8%; Pred. No. 2,1e-06;

Matches 232; Conservative 142; Mismatches 425; Indels 502; Gaps 53;

QY 9 SRLGVPAEPAAELG-----GDASEED-----HPQVCAKCAQFTDPTPE--F 47
Db 40 TRIGVATPKILKGGKRGPFVGDKKRSQVKNNYMVEYYIYNL-GWKLIDATDPKGNW 98
QY 48 LAHON-ACS-----TDPVY-----MVIIGQENPNNSASSEPRPE 82
Db 99 LRYVMNACSGEONLPLEINRAIYYKTLKPIARGEELIWMYNGHDNPRIAAIEBERAS 158
QY 83 G-----HNNPQYMDPE-HSNPPDSGSSVPTDPTMGPERRGESSE 121
Db 159 ASKSSPSKSKKKKSOENKKNKGNKIQIDOLKTSERPFTSANMRDSABPKDEEKP5A 218
QY 122 HPLVATGTAAAGGGGULLASPKIGATPLPPE-STPAPPPPPPPPPGVGSHLINPIL 180
Db 219 SALEQPA-----TLQEVASQEVPELPATPAPAMEQRPDERLEAAACEVNDL 266
QY 181 LEELVLOQROIHOMQTEQICQVLLLSIGQTVGAPASELPGTGTASTKPLLPLE 240
Db 267 GEE-----EEEEEEDEEEBE-----DDDDDELEDEEEFEASMP----- 300
QY 241 SPIKPVQTSKTLASSSSSSSSSGAETPKQAFPHLYHPHLSQHSFASGAGVGSNHPTRAP 300
Db 301 -----NENSVKEPE-----ICDEKP----- 316
QY 301 SPALGSTDOLIASPHLAPPSTTGLLAQOCIGARGLE--ATASPLGLPKPKNGSELSTYG 358
Db 317 -----EDLEEPRTSEET-----LEDCEVYPRAMQIPR--TTEEANG 352
QY 359 EVMGLEPKPGGHNKCFKAKVFGSDSALQILNSH--TGERPYVCNVCNRRPTTRGNIKV 416
Db 353 DVFETMP-----CQHCERKFTTKQGLERHMHIIHSTVNHAFKCKYCGKAFQTQINRR 407
QY 417 HFHRHR---EKYPHVMNHPVPEHLDYVITSSGLPYGMSV-----PP----- 456
Db 408 HERRHEAGLKRKPSQTLP-----SEDL-----ADGKASGENVASKDSSPSLGPDLIMN 459
QY 457 -EKAEEBAATPGGVERKPLVASTTALSTATESLTLTSTAGTAT---APGLPAFNKFLVM 512


```
Db 460 SEKASODT1-----NSSVVEENGVEKELHPCKYCKKVFQHTNMRRHQRYVHERHLIP 512
OY 513 KAVEPRKNKADENTPPSEGSATSGV----- 537
Db 513 KGVRRKGGLEBPAPPAQAOATONVVPSTPEEGEADDDVIMDISNISENLNTYIDG 572
OY 538 -----AESSTATIMQLSKLMTSLPSSMALLTNHFS----- 567
Db 573 KIOTNNNTSNCVDIEMHSASADLYGINCLLT--PVTVEITQNIKTQVPTVEDLPKEPLG 630
OY 568 -----TGSFPL-FLCAALALASPSET-----SK 589
Db 631 STNSEAKKRRATSPALPKIKAEFTSDOPMVPSCSLSLPLSTTTEAVSFHKEKSVYLSK 690
OY 590 LQOLV-----KIDRGAAVATSAASGAPTTSA-----PAPSSSASGSPN 629
Db 691 LKQLOLTOKLTPPAGISATIEIAKLGPVCGANAPAMLPVTSRFRKRTSSPPSPQHP- 749
OY 630 QCVICLRVLSCPRALRLHYGQHGGGERPKCKVCRAFTSGNL--RAHFVGHKASPA-- 684
Db 750 -----ALR-DFGKPSD-----GKAAMTDAGLTTSKSKLSHSDSPAWS 786
OY 685 -----ARAONSCPIC-----QKFT-----NAVTLQOHVBMHGGQIPNGT----- 721
Db 787 LSGRDERETVSPCFDEYKSKKWTASSAFSSVCNQOPLDLSGVKOKAEGTGKTPVQME 846
OY 722 -----ALPEGGAQENGSEOSTVSGAG-----SF 746
Db 847 SVLDLSVHKHKKCSDEKEFKESHVSVOPTCSAVKKRPTTCLMQLKVLNRYGIDLPVEN 906
OY 747 PQOQSOQPSSEBELSEEEEEE-----DEEEDVYTDDESLAGRGESGGE-KAIS 795
Db 907 PADGTSPSPCKSLAOPDPDLGPGSGFPAPTVESYPDVCPSSPALQTPSLISSGLPPL 966
OY 796 VRGD-----SEASGAEEEVGTVAANAATAGKXMDSNEK 828
Db 967 IIPDPSSPPCPVLVATPPRPPLLPYPLRPASSASHPICSPLSNATAOSPLPILSP 1026
OY 829 TTOOSSLPPE--PPDSLDOPQPMEOGSSGLGKEGKPERSSS-----PASALTP-- 879
Db 1027 TVSPSPPIPEVPEPLMSAASGPPPLTSSSSSSSSSSSSSSSSSSSSSPSPPLAISSEV 1086
OY 880 -----EGEATSVLVLELSLOEAM--RKEPGESSSRK-----ACEVGOAAPS 920
Db 1087 SSGDNLASLPMISFKQEBLENEGKPREEPQSAEODVVQETFNKNFVCNVCESPPLS 1146
OY 921 QAALEHQKTHPKEGPLFTCVCFRQGFLEKRLKHHLLAH 961
Db 1147 IKDLTKHLSTHAEMP-FKCEFCVQLFKDKTDLSEHRELLH 1186

RESULT 23
US-09-586-472-4
Sequence 4, Application US/09586472
Patent No. 6323335
GENERAL INFORMATION:
APPLICANT: Huang, Shi
TITLE OF INVENTION: Retinoblastoma Protein - Interacting
Zinc Finger Proteins
NUMBER OF SEQUENCES: 106
CORRESPONDENCE ADDRESS:
ADDRESS: Campbell & Flores LLP
STREET: 4370 La Jolla Village Drive, Suite 700
CITY: San Diego
STATE: California
COUNTRY: USA
ZIP: 92122
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
```

```

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/586,472
FILING DATE: 01-Jun-2000
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 09/528,706
FILING DATE: 17-MAR-2000
APPLICATION NUMBER: US 08/516,859
FILING DATE: 18-AUG-1995
APPLICATION NUMBER: US 08/399,411
FILING DATE: 06-MAR-1995
APPLICATION NUMBER: US 08/292,683
FILING DATE: 18-AUG-1994
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-LJ 4130
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1719 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-09-586-472-4
```

```
Query Match 4.2%; Score 222; DB 4; Length 1719;
Best Local Similarity 17.8%; Pred. 2.1e-06;
Matches 232; Conservative 142; Mismatches 425; Indels 502; Gaps 53;
```

```
OY 9 SRUCVPAGEPRLG-----GDASED-----HPOVCACCAQFTPTPE--F 47
Db 40 TRIGVWATKPIILKGKFGPEVDKRRSQVKNVYMEVYVPL-GWACIDATDPEKGM 98
OY 48 LAHON-ACS-----TDPV-----MYLIGQENPNNSASSEPPE 82
Db 99 LRYVWACSGEONLPLEINRAITYKTLKPIAGBELLVWYNGEDNPETIAALIEERAS 158
OY 83 G-----HNNPQVMDTE-HSNPPDSSSVFTDPTWGPBRKEESSG 121
Db 159 ARSKRSPSKRKKKKKQENKKNKNIQDQLKTSBEDFTSANNRDSAEGRDEKPSA 218
OY 122 HFVYAATGAAGGGGLIASKRLGATPLPE-STAPRPPPPPPPPPGVSGHNLPLI 180
Db 219 SALRQPA-----TLOEVASQEVPELATPAPAMEPOPEPDERLEAAACEVNDL 266
OY 181 LEELRVLOQRQHOMQTEOICQOVLLGSLQTVGAPASPSELPGTGASTKPLPLF 240
Db 267 GEE-----EEEEEDEDEEE-----DDDDDELEDEGEESAAMP----- 300
OY 241 SPIKPVQTSKTIASSSSSSSSSSSGAETPKQAFHLHYPLGSHQHPFSAGVGRSHKTPAP 300
Db 301 -----NENSVKEPE-----IKDEKP----- 316
OY 301 SPALPGSDOLIASPLAPLPSTGTGLAAGCTGAARGE--ATASGLLKPKNNGSELSTYG 358
Db 317 -----EDLLEKPTTSEET-----LEDCEVTPAMQIDP--TKEENAG 352
OY 359 EVMGPLEKPGGRHKCFKCAKVFGSDSALOILNRSH--TGERPYKCNVCGNRPFTTGNLKV 416
Db 353 DYFETFMFP-----COHCERKFTTKQGLERHMHISTVNHAFCKYCGKAKAGTOINRRR 407
OY 417 HFHRR-----EKYPHVQMNPRVPEHLDYVITSSGLPYMSV-----PR----- 456
Db 408 HERRHEAGLKRKRSOTLP--SEDL-----ADKASGEVNAASKDSSPISGLPDCILNN 459
OY 457 -EKAEEAATPGGVERRKPLVASTTALATETSLTLSTAGTAT--APGLPAFNKFLVM 512
Db 460 SEKASODT1-----NSSVVEENGVEKELHPCKYCKKVFQHTNMRRHQRYVHERHLIP 512
```


QY	513	KAVEPKNKADENPPGSEGSISGV-----	537
Db	513	KCVRRKGGLEPPQPAQAQATQVYVVPSTPEEGEGADVYIMDISSINSENLYIDG	572
QY	538	-----AESFATYIMOLSKIMTSLPSWALLTHNFKS-----	567
Db	573	KIOTNNNTSNCVDIEMESASADLYGIMCLLT--PVTVEITONIKITQYVPEEDLDPKPLG	630
QY	568	-----TGSPPL-PLCARALGASSET-----SK	589
Db	631	STNSEAKKRRRTASPPALPKIKAEITDSPPMPFSCSLSTLSTYEAVSFHKERYLSER	690
QY	590	LOOLVE-----KIDROGAVAVYTAASGAPTTSA-----PAESSASACGN	629
Db	691	LKOLLOTODKLTPEPAGISATEIKLAGVYVSAPASMLPVTSRRFRRTSSPPSSQHSB-	749
QY	630	QCVICARLVSCPRALNLYHGQHGERFCKVCYGRATSTKCNL--RAHFNCHKASPA--	684
Db	750	-----ALR-DREKPSD-----GKAWTDAGLTSKKSKLSHSDSPAWS	786
QY	685	-----ARAQNSCPIC-----QKRFET--NAVYLOOHVRAHLGOLIPNGG--	721
Db	787	LSGRDERETVSPCCPEDEYKMSKMTWTASSAPSSVYNQDPDLDISSGVKKQAEGTGTPVOME	846
QY	722	-----ALPBGGAQENGSEDSIVSGAG-----SF	746
Db	847	SVLDSVYHKHCSDSEBGEKFEKESHVQPTCSAVKRRKPTTCMLQKVLLENYNGIDLPIYEN	906
QY	747	FOOQSOQSPPEELISEEEEE-----DEPEEDVYDEDSLACGSHSGE-KAIS	795
Db	907	PADGTRSPSPCKSLEROPPDLPDGGSGFPAFYVESTPYDVCSSPALOTPPSLSSGOLPPLL	966
QY	796	VRGD-----SEBASGAEBEYGTVAATAATGKEMDSNEK	828
Db	967	IPTDSSPPPCBPVLVATPPPLLTPTVPLPAESSASAPHPCCPSPLSNATQSPPLIPSP	1026
QY	829	TTQOSSLPPPP--PPDSLDPOPOMEGSSGVLGKKEGKAPERSSS-----PASALTP--	879
Db	1027	TVSPSPSPPIPEPLMAASFGCPPTLSSSSSSSSSSSSSSSSSSSSSPSPPLISAISVV	1086
QY	880	-----EBGANSVTLVFELSIOEAM--RKEPGESSSRK-----ACEVGGQAAPS	920
Db	1087	SSGDNLEASLPIMISKEOELENEGLTAPREPOGASAEQDVVYOETFNKNFCNVCSPPLS	1146
QY	921	QAALEEHOKTHPKEGELFTVCFCROGELERRATLKKHMLLAH	961
Db	1147	IKDLTKHLSTIAHEMP-FKCEFCVOLFKKOKTJDSJEHFFLH	1186

```

? CLASSIFICATION:
?
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 08/516,859
? FILING DATE:
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 08/292,683
? FILING DATE: 18-AUG-1994
? ATTORNEY/AGENT INFORMATION:
? NAME: Campbell, Cathryn A.
? REGISTRATION NUMBER: 31,815
? REFERENCE/DOCKET NUMBER: P-LJ 1776
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: (619) 535-9001
? TELEFAX: (619) 535-8949
? INFORMATION FOR SEQ ID NO: 4:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 1719 amino acids
? TYPE: amino acid
? TOPOLOGY: linear
? MOLECULE TYPE: protein
? US-09-528-706-4

```

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/528,706
FILING DATE:

OY 568 -----TGSPPL-PLCARALGASPEET-----SK 589
Db 631 STNSEAKKRRATSPALPKIAETDSDPMVPSCSISLPLISITTEAVSFHKEKSVYLSK 690
OY 590 LOOLVE-----KIDROGAVAVMSAAGCAPTTSA-----PAPSSASGPN 629
Db 691 LKQLOQTQDKLTPPAGISATETIAKLGVCVSAFASMLPYTSSFRKRTSSPPSPQHS- 749
OY 630 OCVICLRVLSCPRALRLHYGONGGERPEKCKVCYGFASFSTRGNL--RAHFVGHKASPA-- 684
Db 750 -----ALR-DFGKRPD-----GKAAMTDAGLTKSKSKLESKSDSPAWS 786
OY 685 ----ARAONSCPIC-----OKFT-----NAVTLQOHVRLHGGQIPNGT----- 721
Db 787 LSGRDERETVSPCFDEYKMSKEWTASSAFSSVCNQQPLDLSGKOKABGTGKTPVQWE 846
OY 722 -----ALPEGGAOENGSEOSTVSGAG-----SF 746
Db 847 SVULDLSVHKHCHSSEKGEFESHSHSVQPTCSAVKKRPTTCLQKVLINEXNGIDLPEVN 906
OY 747 PQOQSOQPSPEELSEEEEEE-----DEEEDVTDSDSLAGSGSESGE-KAIS 795
Db 907 PADGTRSPSPCKSLAEQDPDLGPGSGFPAPVVESTPDVCPSPALQTPSLSSGQLPPL 966
OY 796 VRGD-----SEASGAEVEVGTVAATAAGKEMDSNEK 828
Db 967 IPIDPSSPPCPVLYVATPPPLPTVPLPAPSSASBPCHSPUSNAVAQSPPLPILSP 1026
OY 829 TTQOSSLPPPP--PPDSLDQOPMEQSGVLGKKEGKPERSSS-----PASALTP-- 879
Db 1027 TVSPSPPIPTVEVLEMAKSAQCPRLSSSSSSSSSSSSSSSSSSSPRPLSAISSV 1086
OY 880 -----EGEATSVTLVEELSLOEAM--RKEPESSSRK-----ACEVCGQAFPS 920
Db 1087 SSGNDLBSALPLMISFKQBELENEGLKPREEPOSAAEQDVVVOETFNKNFVNCVESPLIS 1146
OY 921 QALLEEHOQKTHPKEGPLFTCVCRGQFLERATLTKHMLLAH 961
Db 1147 IKDLTKHLSTHAEMP-FKCEPCVOLFKDKTDLSEHFLHL 1186

RESULT 25
US-09-262-773-2
; Sequence 2, Application US/09262773
; Patent No. 6225431
; GENERAL INFORMATION:
; APPLICANT: Ballinger, Dennis G.
; APPLICANT: Ding, Wei
; APPLICANT: Wagner, Susanne
; APPLICANT: Hess, Mark A.
; TITLE OF INVENTION: CHROMOSOME 11-LINKED CORONARY HEART DISEASE
; FILE REFERENCE: MYRIAD 3
; CURRENT APPLICATION NUMBER: US/09/262,773
; CURRENT FILING DATE: 1999-03-04
; NUMBER OF SEQ ID NOS: 210
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 648
; TYPE: PRT
; ORGANISM: human
US-09-262-773-2

Query Match 4.2%; Score 219.5; DB 4; Length 648;
Best Local Similarity 19.5%; Pred. No. 8.8e-07;
Matches 136; Conservative 65; Mismatches 263; Indels 233; Gaps 20;

OY 60 VMVITGGENPNSSASBPREGCHNNPQVMDREHNSNPRDSSGVPTDPTWGPBERGEES 119
Db 135 VIVHVGGEVLSEETVNLCAEPESP-ELQDPVQST-----PQSPPEET 178
OY 120 SGHFLVATGTGAAGGGGLILASPKIGATPLPRESTR-----APPPPPPPP 167

Db 179 T-----QSPDLGA---PAEQRPQEEELQTLQESVPEDEPDLF 215
OY 168 PGVSGHLNPLILEELRV-----LQORQHOMQTEQICQVL 206
Db 216 AERSSGDSSEWALLTALSOGLVTFKDVAVCFSDQSDLDLPQTEKEYGEVLEDCGIVV 275
OY 207 LLSLQGTVCAPASPSELPGHGTASTKPLPLFSPRIKPVQTSKTLAASSSSSSSSSGAE 266
Db 276 ---SLSPPIPPDELSOV-----REEEPWPDIDQEPQTEPEILSFYTGDRSKDEE 326
OY 267 TPQOAFNLHYPLRISQHPFSAGVGRSHKPTRAPSPALPGSTQDLIASPHLAFSTTGL 326
Db 327 CLQEDSLDLE-----IHRPV-LGEDEIHQTPDM-----EIVFEDNPGRL 365
OY 327 AAOCLGARGLEATASPEGLKPKNGSGELSYGEVMKPLEPGGRHRCFCAKVCGSDAL 386
Db 366 NERRFQTN-----ISOVNSFVNLRETTPVHPL--LGRHHDCVCGKSTFCNSHL 412
OY 387 QIHRSHTGERPRYKCNVCGNRRFTTGNLKVHFHREKRYHVOVMNPHRVEHLDVITSS 446
Db 413 VRHLRTHGKRPYKMEGCKSTYTRSSHRLARQVHKMNPY----- 453
OY 447 GLPYGMSVPRPKAEDEAATPGGVERKPLVASTTALSAATESLTLSTAGTAPGLPAF 506
Db 454 -----KYLNRKNLEETSPVTOAERTP----- 475
OY 507 NKEYLMKAVERKNKADENTPPGSEGAISGVASSTATLMQSKLMT-----S 554
Db 476 -----SVERPYRCD-----CGKHFRMTSDLRHNOHTHGEKPEFTIGKS 517
OY 555 LPSWALLNHFKSTGSFPLPLCARALGASPEFTSKLOOLEKIDROGAVAVTSAAGAPT 614
Db 518 FSQKSVLTTQRI-----HLGKPYLQEGCEDESEHRYLAHRKTHAAEEL-- 564
OY 615 TSADAPSSASSAGPNOCVYICLRVLSCPRALRLHYGONGGERPEKCKVCYGFASFSTRGNLRA 674
Db 565 -----YLSCGRCGCTHSAFAKHLRGHSAVPRCNCNEGKSFSTRD--- 606
OY 675 HFVGHKASPARAQAONSCPICOKKFTNAVTLQOHVRMH 711
Db 607 HLVRHQRTHGKRPCTCGKSFSGRYHLIRHQRTH 643

RESULT 26
US-09-262-773-4
; Sequence 4, Application US/09262773
; Patent No. 6225431
; GENERAL INFORMATION:
; APPLICANT: Ballinger, Dennis G.
; APPLICANT: Ding, Wei
; APPLICANT: Wagner, Susanne
; APPLICANT: Hess, Mark A.
; TITLE OF INVENTION: CHROMOSOME 11-LINKED CORONARY HEART DISEASE
; FILE REFERENCE: MYRIAD 3
; CURRENT APPLICATION NUMBER: US/09/262,773
; CURRENT FILING DATE: 1999-03-04
; NUMBER OF SEQ ID NOS: 210
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 640
; TYPE: PRT
; ORGANISM: human
US-09-262-773-4

Query Match 4.1%; Score 218.5; DB 4; Length 640;
Best Local Similarity 20.7%; Pred. No. 1e-06;
Matches 129; Conservative 61; Mismatches 246; Indels 187; Gaps 19;

OY 145 LGATPLPPE--STRAPPPPPPPPPGVSGHLNPLILEELRVLQORQHOMQTEQIC 202
Db 144 LGAEPSPNELQDPVOSTPDSQSPDETTQSPDLGA---AEQRPQEEELQTLQESE--- 197


```

OY 203 ROVLLLSIGQVGPASPSELPGTGTASTKPLPLFSPDKPYQTSKTIASSSSSSSS 262
Db 198 -----VPEDEP-DLPAERSSSDSEWALITLSSQGLVTRDAVVCSDQDMS 244
OY 263 SGAEPTKPAEFHLVHPLGSOHPFSAGVGRSHKPTAP-----SPALPGSTD-OLI 312
Db 245 DLDPOTKE--FYGEVLBE---DCGIYVLSLFPPIPRPDEISQVREDEPWPDQEPET 298
OY 313 ASPHLAPSTTGLA---AQCIGA-RGLEATASPGLKP---KNGSGELSYGEVMPLE 365
Db 299 QEPFLLSFTYIGDRSKDEECLQEDLSLSDIHVPYLGEPFIHQTPDMEIVFEENPGRLN 358
OY 366 KP-----GGRHKCRFCAKVYFSDSALOIHRSNNGERYK 400
Db 359 ERFRGTNISQVNSFVNLRETTPVHPLLGRHHCSCVCKSTCSHLVRLKHTGEKRY 418
OY 401 CNVGNRPFTTNGNLKVHFRHREKYPVOMNPPVPEHLDYVITSSGLPYGMSVPERKAE 460
Db 419 CMCGKRYTSSSHLARKQVKHKNAPY-----KYPILNRKN 453
OY 461 EEAATGSGVERKPLVASTTALSATESLILSTASGATATAPGLPAPNKVFLMKVPERKN 520
Db 454 LEETSYTQAEKRP-----SVERPYR 474
OY 521 ADENTEGSGSALSGVAESSTATLMOLSKLMT-----SLPSWALLTNHFKST 568
Db 475 CDD-----CGKHFRWTSIDLVRHORHTGKPFCTTCKGSPSQSVLTHHRI 522
OY 569 GSPPLLCARALGASPETSLOQLVEKIDRGAYAVTSAASAPPTTAPAPSSASSGP 628
Db 523 -----HLGKPYLGCGEDEFSHRRLAHRKTHAAEEL----- 556
OY 629 NOCVIDLRYLSCPRLRLHYOHGGERPFCKVCGRAFSTRGNLRAHFVGHKASPARAQ 688
Db 557 YLCSGGRCTHSAFAKHLGHNASVRCRCNECKSFSKRD---HLVRHQHTHGEKP 612
OY 689 NSCPTQCKKTNVTLQOHVYRM 711
Db 613 FTCPCTGKSFSGRYHLTRHORTH 635

RESULT 27
US-08-102-942A-4
Sequence 4, Application US/08102942A
Patent No. 5726288
GENERAL INFORMATION:
APPLICANT: Call, Katherine M.
APPLICANT: Glaser, Thomas M.
APPLICANT: Ito, Caryn Y.
APPLICANT: Buckler, Alan J.
APPLICANT: Pelletier, Jerry
APPLICANT: Haber, Daniel A.
APPLICANT: Rose, Elise A.
APPLICANT: Housman, David E.
APPLICANT: Bruening, Wendy
APPLICANT: Darveau, Andre
TITLE OF INVENTION: Localization and Characterization of the
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESS: Hamilton, Brook, Smith & Reynolds, P.C.
STREET: Two Militia Drive
CITY: Lexington
STATE: Massachusetts
COUNTRY: U.S.
ZIP: 02173
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:

```

```

APPLICATION NUMBER: US/08/102,942A
FILING DATE: 02-AUG-1993
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Granahan, Patricia
REGISTRATION NUMBER: 32,227
REFERENCE/DOCKET NUMBER: MIT-5194A2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-861-6240
TELEFAX: 617-861-9540
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 449 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-102-942A-4

Query Match 41%; Score 215.5; DB 1; Length 449;
Best Local Similarity 25.8%; Pred. No. 1e-06;
Matches 101; Conservative 43; Mismatches 130; Indels 117; Gaps 23;

OY 124 LVATGTAGGGGGLILASPKLGA---TPL-----PEST-----PAPPPPPPPPG 169
Db 11 LPAVSSLGGGGGGGGL--PVSGARQWAPVLDFAEPGASAYSLGFPAPPPAPPPPPPP 67
OY 170 VGSGLNIPILIEELRYLQKROIHOMQTEQICROYLLLSLQ---TVGA-----PA 219
Db 68 -----PSHFKEPSSWGAEPHEEQLSAFTLHFSQFTGTACACRAGPGRP 115
OY 220 SPSEL-PGTGASTKPLPLFPSPKPY---OTSKL-----ASSSSSSSSSGAETPKQA 271
Db 116 PPGQASSGQARMFNAPNAPYLPSCLESQPTIRNOGISTVITDGAISYCHTSHHAAQFPNHS 175
OY 272 FPHLYHPLGSOHPFSAGVGRSHKPTP-----APSPALPGSTDOLIASPULA---FPS 321
Db 176 FKH-EDPMGQ---SGLGQYQSVPPVYCGHPPTDCTGQALLLRPPSSDNLQYM 229
OY 322 TTGLLAQC-----IGAA-RGLEATASPGLL---KKNSGELS----- 356
Db 230 TSQL---ECMTWNOMNIGATLKGMAAGSSSVKWTGEGSNHGTGYSENHTAPILCAQY 286
OY 357 ---YGEVWG--PLEKPG-----RHKCRF--CAKVFSGDSALOIHLS 392
Db 287 RIHTGVERGIQDVRRVSGVAPTLVRSASSTSEKRPFKATYGCCKNRYKLSHLQHSRK 346
OY 393 HGERPYKAV--CGNFTTGNLKVHFRH 421
Db 347 HGEKPYQCDPKDCERRFSRSDQLKRHQRH 377

RESULT 28
US-09-037-179B-4
Sequence 4, Application US/09037179B
Patent No. 6316599
GENERAL INFORMATION:
APPLICANT: Call, Katherine M.
APPLICANT: Glaser, Thomas M.
APPLICANT: Ito, Caryn Y.
APPLICANT: Buckler, Alan J.
APPLICANT: Pelletier, Jerry
APPLICANT: Haber, Daniel A.
APPLICANT: Rose, Elise A.
APPLICANT: Housman, David E.
APPLICANT: Bruening, Wendy
APPLICANT: Darveau, Andre
TITLE OF INVENTION: Localization and Characterization of the
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESS: Hamilton, Brook, Smith & Reynolds, P.C.
STREET: Two Militia Drive
CITY: Lexington
STATE: Massachusetts
COUNTRY: U.S.
ZIP: 02173
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:

```



```

: PRIOR APPLICATION NUMBER: US 07/614,161
: PRIOR FILING DATE: 1990-11-13
: PRIOR APPLICATION NUMBER: US 07/435,780
: PRIOR FILING DATE: 1989-11-13
: PRIOR APPLICATION NUMBER: US 07/795,323
: PRIOR FILING DATE: 1994-09-27
: NUMBER OF SEQ ID NOS: 21
: SOFTWARE: FASTSEQ for Windows Version 4.0
: SEQ ID NO 4
: LENGTH: 449
: TYPE: PRT
: ORGANISM: Unknown
: FEATURE:
: OTHER INFORMATION: Mutine
US-09-037-179B-4

Query Match      4.1%; Score 215.5; DB 4; Length 449;
Best Local Similarity 25.8%; Pred. No. 1e-06;
Matches 101; Conservatve 43; Mismatches 130; Indels 117; Gaps 23;

QY 124 LVNATGTAAGGGGGLILASPKLGA---TPL-----PPESF-----PAPPPPPPPPPG 169
    | : | : | | | | | | | | | | : | : | : | : | : | : | : | : | : |
Db 11 LPRVAVSSLGGGGGGCGCL--PVSGARQMAPVLDPAPGASAVGSLGFPAPPPPPPP 67
    | : | : | | | | | | | | | | : | : | : | : | : | : | : | : | : |
QY 170 VGSGHILPILLEBLRYLQQRQIHOMQMTQICRQVLLGSLGQ---TVGA-----PA 219
    | : | : | | | | | | | | | | : | : | : | : | : | : | : | : | : |
Db 68 -----PHSFIKQEPSSWGGAERHEEQCLSAFTLHFSQFTGTAGCRGFGFRP 115
    | : | : | | | | | | | | | | : | : | : | : | : | : | : | : | : |
QY 220 SPEEL-PTGTASTTKPLPLPSPIKPV---QTSKTL---ASSSSSSSSSGAETPKQA 271
    | : | : | | | | | | | | | | : | : | : | : | : | : | : | : | : |
Db 116 PRSQASSGQARMFENAPRLPSCLESQPTIRNOGYSTVFEDGASVYGHGTPSHNAQFPMHS 175
    | : | : | | | | | | | | | | : | : | : | : | : | : | : | : | : |
QY 272 FFLYHPLHLSGSHPRVAGVGRSHKPTP-----APSPALPGSTDOILASPHLA---FPS 321
    | : | : | | | | | | | | | | : | : | : | : | : | : | : | : | : |
Db 176 FKH-EDRMGQD-----GSLGEQYSVPRPVYGCHTPTDSTGSOALLLRTYSSDNLXQM 229
    | : | : | | | | | | | | | | : | : | : | : | : | : | : | : | : |
QY 322 TTGLLAAC-----LGAA-RGLENTASPGLL---KPRNGSELS----- 356
    | : | : | | | | | | | | | | : | : | : | : | : | : | : | : | : |
Db 230 TSQL---ECMTWNOMNLGATLIGMAAGSSSSSVKWTGOSNHGTGEENHTAPILCGAQY 286
    | : | : | | | | | | | | | | : | : | : | : | : | : | : | : | : |
QY 357 -----YGEVMG--PLEKPG-----NNKCRF--CAKYFGSDSALQIHRS 392
    | : | : | | | | | | | | | | : | : | : | : | : | : | : | : | : |
Db 287 RIHTHGVRFGIQDVRVRSVAVPTLVRSASSETSEKRPKCAIPGCKNKRYFKLSHLMHSRK 346
    | : | : | | | | | | | | | | : | : | : | : | : | : | : | : | : |
QY 393 HTGERPYKCNV--CGNRETTTGNLKVYHFRH 421
    | | | | | : | : | : | | | | | |
Db 347 HTGEKPYQCDFKDCERRRFSRSDQLKRNQRRH 377
    | | | | | : | : | : | | | | | |

RESULT 29
US-08-317-310A-15
: Sequence 15, Application US/0317310A
: Patent No. 5858701
: GENERAL INFORMATION:
: APPLICANT: WHITE, MORRIS R
: APPLICANT: SUN, Xiao Jian
: APPLICANT: PIERCE, Jacalyn H.
: TITLE OF INVENTION: THE IRS FAMILY OF GENES
: NUMBER OF SEQUENCES: 64
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: LAHIVE & COCKFIELD
: STREET: 28 State Street
: CITY: Boston
: STATE: Massachusetts
: COUNTRY: USA
: ZIP: 02109
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: ASCII text
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/317,310A

```

```

FILED: 03-OCT-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Louis Myers
REGISTRATION NUMBER: 35,965
REFERENCE/DOCKET NUMBER: JDP-022
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)227-7400
TELEFAX: (617)227-5941
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 1234 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
FRAGMENT TYPE: Internal
US-08-317-310A-15

Query Match 4.1%; Score 215; DB 2; Length 1234;
Best Local Similarity 20.4%; Pred. No. 4e-06;
Matches 216; Conservative 111; Mismatches 358; Indels 372; Gaps 55;

QY 2 AHESERSRLGVPRGEPRLGDSAEEDHPRVCAKCAQFTDTEFLAHQNA--CSDPP 59
||| ||||| : : : : : ||
DB 351 AHRHRSRLPLPPNHSRSL-----PMDSRCSFSPATSVLSUSSSTSGHGSYDC 401
60 VMTVIGQENNNNSASSEPREGHNNPQMDTEH--SNPPDSSS----VPTDPTWGER 114
402 LF-----PRSSASVSGSPS--DGGTISDEYGVSSPCDFRFSFVSVPDLSGHTPPA 451
QY 115 RGEES-----SGHFLVAATGTAAAGGGGGLLASKLCAATPLPP----- 152
||| : : : : : : : : : : : : : : :
DB 452 RGEELSNYICMGKGAFTLAAPNGHYTL----SRGNGHRYIPGANLGPALPGDEA 506
153 -----ESTRAPPPP-----PRPPPCVAGS----- 173
DB 507 AGAADLNNFRKPKTHSAGTSPTISHOKTSSQSVASIEETEMPAAYPRGGSGGRLP 566
QY 174 -----HLNIPLEELRVLAQROIHQMONTEDICROVLLLSLGQTVGADAPSEL 224
567 YRHSAFVPTH--SYPEGLEMHNHLERGGHNRPDTSMLHTDD---GYPMSPGVAPVPSNR 622
QY 225 PGGTASTKPLP-----LFSPIK--PVQT-----SKLASSSSSSS 260
623 KGNQ--DYMPSPKVSAPAQDITNPIRHPQVDPNGVMMSPSGSCSPDIGGSSSSS 679
DB 261 SSSGAETPKQAFHLYHPLGSOH--PFSAGVGRSHKPTPAPSPALPGSTDLAAPHLA 318
680 SISAA-----PGSSSYGKRWI--NGVGGH-----THALPRAKPRVEGGGKL 720
QY 319 FPSTTGLIAOCLGARGLLEATASPGLLKPKNGSGELSYDEVWGPLEKP----- 367
721 LPTGIDYMNMSFVGSNN--TSSP-----SECYUG--EDPQHKPVLSTYSLPSRF 766
368 -----GGRKCRCAKVFSGDSALQIHLSHTGEPYKCNV----- 403
DB 767 KHTQRPGEPEBARIQ-----HLRLSSSGRLRYATATADSSSTSSDLSL 812
QY 404 --CGNRETTGNLKVHFHRRHREKYPHVOMNPHRPHLDVY-----TSSGLP 449
||| : : : : : : : : : : : : : : :
DB 813 GYGCGAR--PESSL--THPHNN-----VLQPH--LPKKVLTAAOTNSRLAPTRLSTLDP 861
450 YGMSVPRPEKAE-----EAATPG-----GGVERKPLVASTALSTESTLILS 492
DB 862 KASTLPVREQQOQOOSLSHPPEKSPGEVYNIETFGSGQPEYLAGPATSRSSPS----VR 917
QY 493 TSAGTATAPGLAFNKFVLMKAVEPKKNKADENTPQSEGSASISGVAESSTATLMOLSKIM 552
918 CPQGLHAPARETSGSEBYNMMDLGPGRRAATWQEGGVELGRI--GRAPGSAIVCPRT-- 974
QY 553 TSLPFWALITNFKSYSGSPPLP--CARALGASPSSTKLOOLYEKIDROGAVANTASMS 610
||| : : : : : : : : : : : : : : :
DB 975 -SVNP-----SRGDDVNTMOIGCR-----QSVTVDSVPAP--VSAYAMRGIAAEKKSIL 1020

```


GENERAL INFORMATION:
APPLICANT: Herlyn, Meenhard
APPLICANT: Morris, Jennifer
APPLICANT: Rauscher III, Frank J.
APPLICANT: Rodeck, Ulrich
TITLE OF INVENTION: WTI Monoclonal Antibodies and Methods of
TITLE OF INVENTION: Use Therefor
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: Howson and Howson
STREET: Spring House Corporate Cntr, PO Box 457
CITY: Spring House
STATE: Pennsylvania
COUNTRY: USA
ZIP: 19477
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/234,783
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Bak, Mary E.
REGISTRATION NUMBER: 31,215
REFERENCE/DOCKET NUMBER: WST480UA
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-540-9200
TELEFAX: 215-540-5818
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 429 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-234-783-4

Query Match 4.1%; Score 214.5; DB 1; Length 429;
Best Local Similarity 24.3%; Pred. No. 1.1e-06;
Matches 91; Conservative 43; Mismatches 122; Indels 119; Gaps 18;

133 GGGGGLILASPKLGATP-----LPPEST-----PAPPPPPPPPPGVSGHLNIP 178
19 GGGGGCAL--PYSGAQMAYLDFAFPGASAYGSLGCPAPPPAPPPPPPHS----- 70
179 LLEELRYLQROQHOMOMTEQICROVLLGSLGQ--TVGA-----PASPSEL-PGT 227
71 -----FIKQPSWGAERHEQCLSAFTVHFGQFTGTAGACRYGPRGPPPSQASSGQ 124
228 GTASTKPLPLFSPRIKRYQVTSKTLASSSSSSSSSSG-----AETPKQAFHLYHPLG 280
125 ARMPFNAPYLPSCLESOPAIRNOGYSTVTFDGTSPSYGHTPSHNAQFPHNSFKH-EDPMG 183
281 SQHPFSAGVGNHKTPT-----APSPALPGSTQDLIASPHLA---FPSTTGL----- 325
184 QQ-----GSLGEQYQVPRPVYGCNTPDSCSTGSQALLKRTYSSDNLQMTSOLECMTW 238
326 ---LAAOCLGAARGLAAT-----ASPGILKPKN 350
239 NQMNLGATLKGHSTGVESDNHTTPIILCGAQRHITHGVFRGIQDVRRVGVAPTLVRSAS 298
351 GSGELSYGEVMGPLEKPGGRHKCRF--CAKYFGSDSALOILHRSHTGERPYKCNV--CGN 406
299 ETSE-----KRP---FMCAYPGCKNRYFKLSHLQMSRKHTGKRPYQCDFKDCER 345
407 RETTGNLKVHPRH 421
346 RFSRDQLKRRHRH 360

US-08-456-907-4
; Sequence 4, Application US/08456907
; Patent No. 5633142
GENERAL INFORMATION:
APPLICANT: Herlyn, Meenhard
APPLICANT: Morris, Jennifer
APPLICANT: Rauscher III, Frank J.
APPLICANT: Rodeck, Ulrich
TITLE OF INVENTION: WTI Monoclonal Antibodies and Methods of
TITLE OF INVENTION: Use Therefor
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: Howson and Howson
STREET: Spring House Corporate Cntr, PO Box 457
CITY: Spring House
STATE: Pennsylvania
COUNTRY: USA
ZIP: 19477
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/456,907
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/234,783
FILING DATE: 28-APR-1994
ATTORNEY/AGENT INFORMATION:
NAME: Bak, Mary E.
REGISTRATION NUMBER: 31,215
REFERENCE/DOCKET NUMBER: WST480UA
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-540-9200
TELEFAX: 215-540-5818
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 429 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-456-907-4

Query Match 4.1%; Score 214.5; DB 1; Length 429;
Best Local Similarity 24.3%; Pred. No. 1.1e-06;
Matches 91; Conservative 43; Mismatches 122; Indels 119; Gaps 18;

133 GGGGGLILASPKLGATP-----LPPEST-----PAPPPPPPPPPGVSGHLNIP 178
19 GGGGGCAL--PYSGAQMAYLDFAFPGASAYGSLGCPAPPPAPPPPPPHS----- 70
179 LLEELRYLQROQHOMOMTEQICROVLLGSLGQ--TVGA-----PASPSEL-PGT 227
71 -----FIKQPSWGAERHEQCLSAFTVHFGQFTGTAGACRYGPRGPPPSQASSGQ 124
228 GTASTKPLPLFSPRIKRYQVTSKTLASSSSSSSSSSG-----AETPKQAFHLYHPLG 280
125 ARMPFNAPYLPSCLESOPAIRNOGYSTVTFDGTSPSYGHTPSHNAQFPHNSFKH-EDPMG 183
281 SQHPFSAGVGNHKTPT-----APSPALPGSTQDLIASPHLA---FPSTTGL----- 325
184 QQ-----GSLGEQYQVPRPVYGCNTPDSCSTGSQALLKRTYSSDNLQMTSOLECMTW 238
326 ---LAAOCLGAARGLAAT-----ASPGILKPKN 350
239 NQMNLGATLKGHSTGVESDNHTTPIILCGAQRHITHGVFRGIQDVRRVGVAPTLVRSAS 298
351 GSGELSYGEVMGPLEKPGGRHKCRF--CAKYFGSDSALOILHRSHTGERPYKCNV--CGN 406
299 ETSE-----KRP---FMCAYPGCKNRYFKLSHLQMSRKHTGKRPYQCDFKDCER 345

[illegible]

```

OY      306  TGSFPLKRALUASPSFISRLQDLVERIDKQGAVAATISASGAPTTAPAPSSSSASG  627
      |-----PHANGIALK-----EORAYEVLRAAS---ENSODAFRYVSTSG  366
OY      628  PN----QVCICLRVLSCPRAL-RLH---YGOHGGEPFCKYCG-----RAESTRG  670
      |-----|-----|-----|-----|-----|-----|-----|-----|
Db      367  EOLKYKCEHC-RVLFIDHWYTIHMGCHGCGHFRDPECNMCGYHSQDRYEFSSHTTG  425
OY      671  NLRAH  675
      |-----|-----|-----|-----|-----|-----|-----|-----|
Db      426  EHRVH  430

RESULT 38
US-09-110-517-2
; Sequence 2, Application US/09110517A
; Patent No. 6248520
; GENERAL INFORMATION:
; APPLICANT: Roeder, Robert G
; APPLICANT: Pondell, Joseph D
; APPLICANT: Yuan, Chao X
; APPLICANT: Ito, Mitsuhiro
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING NUCLEAR HORMONE
; FILE REFERENCE: 600-1-224
; CURRENT APPLICATION NUMBER: US/09/110,517A
; CURRENT FILING DATE: 1998-07-06
; NUMBER OF SEQ ID NOS: 51
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 1581
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-110-517-2

Query Match          4.0%; Score 209.5; DB 4; Length 1581;
Best Local Similarity 19.3%; Pred. No. 1.3e-05;
Matches 174; Conservative 91; Mismatches 263; Indels 373; Gaps 33;

5  SERSRLGAPGEPALGGDASEEDHPQVCAKCAQAFDTPTPEFLAHQNMKSTDPVWVII 64

```


Db 900 SOALNTLGVRM-----LGDNCE-----TFKGNQADYDFSIISVA 937
QY 65 G-----GOENP-----NNSASSEPRPEGHNP----- 87
Db 938 GKALAPADIMENHSGSGLTLTGDLCKEKTQKRYKEGNNGTNSITLSCGLDLSKPGKRSR 997
QY 88 -----QVMDHEHNPDSGSSVP--TDPTWGPERRCEESSGHFLVATGTA 131
Db 998 TRPNDSKSKDKPRKRRKADTEGSPSHSSNRPTPTP-----STG-----GSK 1041
QY 132 AGGGGGLILASPKLGTATPLPESTPAPRRPPRRPPCVGSGHLNPLILEELRVLOQR 191
Db 1042 SPGSMG-----RSQTPRPVATPTPT----- 1061
QY 192 IHOMQTEGICROVLLIGSLGQYVAPASPELPGTASSTKPLPLPLSPKPVQTSKT 251
Db 1062 ----KTTIPIKGTVM-----VGKPSHGOYTSGSVSSGSS-----KSHHS 1099
QY 252 LASSSSSSSSGAETPKQAFPHLYHPLGSHQHPGAGVGRSHKPTPAPSPALPGSTDL 311
Db 1100 HSSSSSSSSASTSKMKSSKSSGSSSKLSSMYSOGSSGSSOSKSSSGGKPGSS--- 1156
QY 312 IASPHLAPSTGTLLAQCIGAARGLEATASPLLKPKNGSELXYGEVGP-----LE 365
Db 1157 -----PIT-----KHGLSSGSSSTKMKPGKPSLMPRLSKPNTSPSHR 1197
QY 366 KKGGRKRCRCACAVFGSDSLQIHLRSHGTGERPYKCNVCNRTTGNLKVHHRHREKY 425
Db 1198 PEGGSDKL-----ASPMK----- 1210
QY 426 PHVQMPNRPVREHLDYITSSGLPYGMSVPRPEKAEPEATPGGVERKPLVASTALSAT 485
Db 1211 -----PVP-----GTP-----PSKAKSPISGSG-----SHMSGT 1237
QY 486 ESLLTLSTAGTAPGLPAPNKVLMKAVEPKNADEMTPEGSESAISGVAESSTATL 545
Db 1238 SSSSGMKSSGSLGSSGL-----SOKTPPSSNCTASASSPSSSGSS 1279
QY 546 MQLSKMLTSLPSMALLNNHFKSTGSFPLRCALCALGASPESTKLOOLVKEKDPOGAVAV 605
Db 1280 MSSSQ-----NOHGSSKGK-----SPSRNKK-PSLTAVIDLKHGCV 1315
QY 606 TSAASGA-PTTSAPAPASSASSGPRQCVCILRVLSQPRALRLHYGQHGRRPFCKYGR 664
Db 1316 TSGPGEEDPLDGOGVSTNSSHP-----MSSHNMSGGE--FGK---R 1355
QY 665 AFSTRGMLRAHFVGHKASPAARAQNS-----CPICOKKFTNAVTLQ 705
Db 1356 EKSDKDKSVSTSSSVSSKKTSESKNVGSTSVAKIITSKHGGSPSIKAK-----VTLQ 1411
QY 706 QHYRMHLGGQIP-----NGSTALPREGGAQENGSEQSVSGA-----GSFP 747
Db 1412 KPGESSGEGLRQVMASSKNYSPPLTSGTSPKHGRGSPSHSKSPAYTPQONDSESESGSI 1471
QY 748 QOQSQQSPPEE-----ELSEEEDEDEDEEDYTDSDSLAGRSESGEKAISVBCD 799
Db 1472 AEXSYQSPSSDDPLRLPRLYSTEKKHKKKKKKKKYKDKORDRDRODKKKHSHIPE 1531
QY 800 S 800
Db 1532 S 1532

RESULT 39
US-08-711-417C-196
; Sequence 196, Application US/08711417C
; Patent No. 6228611
; GENERAL INFORMATION
; APPLICANT: Georgopoulos, Katia A.
; TITLE OF INVENTION: IKAROS: A T CELL PATHWAY REGULATORY GENE
; NUMBER OF SEQUENCES: 202
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.

STREET: 225 Franklin Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows 95
SOFTWARE: FastSeq for Windows Version 2.0b
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/711,417C
FILING DATE: 05-Sep-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/238,212
FILING DATE: 02-MAY-1994
APPLICATION NUMBER: 08/121,438
FILING DATE: 14-SEP-1993
APPLICATION NUMBER: 07/946,233
FILING DATE: 14-SEP-1992
ATTORNEY/AGENT INFORMATION:
NAME: Myers, Louis P.
REGISTRATION NUMBER: 35,965
REFERENCE/DOCKET NUMBER: 10287/007001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617/542-5070
TELEFAX: 617/542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 196:
SEQUENCE CHARACTERISTICS:
LENGTH: 461 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
FRAGMENT TYPE: internal
SEQUENCE DESCRIPTION: SEQ ID NO: 196:
US-08-711-417C-196

Query Match 3.9% Score 206.5; DB 4; Length 461;
Best Local Similarity 23.5%; Pred. No. 4,1e-06;
Matches 110; Conservative 44; Mismatches 162; Indels 153; Gaps 21;

QY 329 QCLGAARGLEATASPLLKPKNS-----GELSYGEVMPLEKPGGRHKRCFCACAVFGSDS 384
Db 22 ECAGDRLMLDASG-----EKMNGSHRDQSSALSGLV-GGIRLPNGKXKDCIGIICIPN 75
QY 385 ALQIHLNHSHTGERPYKCNVCGRFTTRGNLKVHFRHREKY----- 426
Db 76 VLMVHKRSHGTGERPQCNOGASFQKGNLRIHKLHSGKPFKCHLCNVACRRRDALTG 135
QY 427 ----HVQMPNPH-----PVREHLD---YVITSSGLPYGMSVPRPEKAEPEATP 466
Db 136 HLRHNSVKKPRHKCGYGRYKQRTSLDEHKRECHNYLESMDLP-GTLXLPVKEELKHSEM 194
QY 467 GG-----GVERKPL---VASTALATESITLLTSTAGTAP-GLPAENKFLVLMKAVEP 517
Db 195 AEDLCKTISERSLVLDRLASNVAKRKSSMPQKFLGDKGLSDTPYDSATYERKENEMKSHV 254
QY 518 KKKADEN-----TPPG-----SEGSAISGVAESSTAT- 544
Db 255 MDOAINNAINYLGAESLRPLVOTPPGGESEVPIVSPMYQLHRRSEGTFRSHNSAODSAVE 314
QY 545 ---LMQLSKLMS---LPS-----WALLNNHFKSTGSFPLRCAR 578
Db 315 YLLLSKRAKLVSERKAPSSNSQDSTDTESNNDEQKSGLLYLLNH-----IARR 364
QY 579 ALGASPESTSKLOOLVEKI--DRQGAVAVTSAAAGADPTSPADSSASSGPMQVCITLR 636
Db 365 AQRVSLKEHRAVLDLLRAASENSQDALRVYS-----TSGQMKVYKCEHC-R 410
QY 637 VLSCPRL-RLHYGQHGGERPFKCYG-----RAFSTRGNLRAH 675
Db 411 VLFIDHVAVTYTHMGCHGFRDPFECNMCGYHSQDRYEFSSHTTREGHRH 459


```

RESULT 40
US-08-046-585-5
Sequence 5, Application US/08046585
Patent No. 545362
GENERAL INFORMATION:
APPLICANT: Lamarco, Kelly
APPLICANT: Wilson, Angus
APPLICANT: Herr, Winship
TITLE OF INVENTION: A NOVEL EUKARYOTIC TRANSCRIPTION PROTEIN:
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSER: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
STREET: 4 Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/046,585
FILING DATE: 12-APR-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Osman, Richard A
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: A-57503-1/RAO
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
TELEX: 910 277299
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 2035 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-046-585-5

Query Match 3.9%; Score 206.5; DB 1; Length 2035;
Best Local Similarity 19.9%; Pred. No. 2, Be-05;
Matches 193; Conservative 94; Mismatches 374; Indels 311; Gaps 40;

QY 125 VAATGTAAGGGGGLIASPKL-----GATPLPEESTPAPPPPPPPPGVSGHLNI 177
DB 803 VMTSGT--GAPAKITAVPKIATGNGGQGVTVVILKAPGPGITLRTVP--MGVRLVT 858
QY 178 PLILEELRYLQORIHOMQMTEDQICROVLLGSLGOTVGAAPSPSELPGTGAASKPL 237
DB 859 PVTYSAPK-----PAVTLVVKGTGTGVTGTVGTVS--TSLAGAGGHSSTASLA 907
QY 238 PLFSPDKPQVOTSKTLAS--SSSSSSSSGAEPTKQAFNLHYPLGSOHPFSAGVGRSHK 295
DB 908 --PITTLGTLATLSSOYINPALTIVSAOQTITLTAAGGLTPPTTTPQVPS-----Q 956
QY 296 PTPAPSPALPGSTD-----QLIASPHLAPSTTGLLAQCLGAARGLEATASPGIL 346
DB 957 PTOVTLTLPAPSGEVAQRPVHDLPVYSILASPTTEOPTATVTIA----- 997
QY 347 KPKNGSGELSGEYVNGPLKPGGRHKCRFCAPKVGSGSALQIHLRSHTGERTPKYKVCN 406
DB 998 --DSGGGVQVGETVTLVCSNP-----PCETHETGT 1025
QY 407 RFTTRGNLKVHFHRRKKYRHYQMNPRPVREHLDYV-----ITSS--GLPYGMSV- 454
DB 1026 TNTATTVVAN-----LGGHPRPTQVQFVCDRQEAASLVYTSIVGQGNQSVYR 1073

```

```

QY 455 -----PPEKAEAEATPG-----GVEKPLVASTTALSTESTLTLSTAG--- 496
DB 1074 VCSNPCEHETGTNTATATTSNAGOHGCSNPCEHETGTNTATATTSNAGANOR 1133
QY 497 ---TATAPGLPAFNKE--VLMAVEPKNKADENTPGESEGAISGVASESTATLMQLSKIM 552
DB 1134 DARRACAAGTPAVIRISVATGALF-----AAQSKSCCQTRQTSATSTTMTVMA 1182
QY 553 TSLPSSMALTNHFKSTGSPFLPLCARALGASPSETSKLOLVEKIDROGAVAVTSAASG- 611
DB 1183 TGAPC-----SAGLPLGSPMARPEGGRSPAFVQLAPLSKV-RLSSPSIKDPAGR 1232
QY 612 ---APTSAAPASSASSSGPNQVC-----LRLVSCPALRLHYQHG 652
DB 1233 HSHAIVSTAAATRSSVGAGEPRMAPVCESLQGGSPSTTVTVALCALCPSAT----- 1284
QY 653 GERPFCKKVCGRAFSTGNLRAHFVG----HKASPARAQNQC--PICQKFT---NAV 702
DB 1285 -----VQVC-----SNPCEHETGTNTATTSNAGSAQRYCSNPCEHETGTHTAT 1334
QY 703 TLOOHVAMHLGGQILPNCGTALPEGG-----GAQENGSEOST-----VSGAGSFPOQOQ 752
DB 1335 TATS-----NGTGOPEGGQGPAGRPCEHTQTTSTGTTMSVSGALLPDATSS 1383
QY 753 Q-----PSPEELSEEEDEEDEDVDEDSLGRGSE 787
DB 1384 HRTVESGLEVAAPSVTPQACTALLAPFTTORVCNSNPCEHETGTHTATVYTSNMSN 1443
QY 788 SGGEKAI SVRGDSEEAAGAEVEGTVAATAAGKEMDSNEKTTOOSLPP-----PPDS 843
DB 1444 ODPPPAASDQGEVESTGDSVNITSSSAITTVSSTLFRVAVTVTQSTVPQPSVPPPEE 1503
QY 844 LD-QPQPMQ-----GSSVILGKEE---GKKEP-----RSSSPASALTPEG 881
DB 1504 LOVSPGPRQQLPPRLQLQASATLMSGAEVLASQTELPAAVDLSSTGPESSGQESAG 1563
QY 882 EATSVTL-----VEELSLQEMARKE-----PGESSSRKACEVCGQ 916
DB 1564 SAVVATVVVQPPPTQSEVQSLPQELMARAQCTTTLMTVGLPELAVYTAAMAQ 1623
QY 917 APPSQALEEHQ 928
DB 1624 A-----AATEBAQ 1631

```

Search completed: January 13, 2003, 15:23:14
Job time : 53.0224 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 13, 2003, 15:19:10 : Search time 10.0149 Seconds
(without alignments)
1946.896 Million cell updates/sec

Title: US-09-988-117-1
Perfect score: 5277
Sequence: 1 MAHESRSRRUGVPAEPAE.....PSTTGTGLSPRRKDDPPIIP 1005

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 118974 seqs, 19401057 residues
Total number of hits satisfying chosen parameters: 118974

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database : Published Applications_AA:*
1: /cgn2_6/ptcodata/1/pubpaa/US08_NEM_PUB.pep:*
2: /cgn2_6/ptcodata/1/pubpaa/PCIT_NEW_PUB.pep:*
3: /cgn2_6/ptcodata/1/pubpaa/US06_NEM_PUB.pep:*
4: /cgn2_6/ptcodata/1/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptcodata/1/pubpaa/US07_NEM_PUB.pep:*
6: /cgn2_6/ptcodata/1/pubpaa/US07_PUBCOMB.pep:*
7: /cgn2_6/ptcodata/1/pubpaa/PCITUS_PUBCOMB.pep:*
8: /cgn2_6/ptcodata/1/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptcodata/1/pubpaa/US09_NEM_PUB.pep:*
10: /cgn2_6/ptcodata/1/pubpaa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptcodata/1/pubpaa/US10_NEM_PUB.pep:*
12: /cgn2_6/ptcodata/1/pubpaa/US10_PUBCOMB.pep:*
13: /cgn2_6/ptcodata/1/pubpaa/US60_NEM_PUB.pep:*
14: /cgn2_6/ptcodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	5277	100.0	1005	US-09-988-117-1	Sequence 1, Appl1
2	5277	100.0	1005	US-09-812-471-1	Sequence 1, Appl1
3	5277	100.0	1005	US-09-812-633-1	Sequence 1, Appl1
4	4677.5	88.6	1002	US-09-988-117-3	Sequence 3, Appl1
5	4677.5	88.6	1002	US-09-812-471-3	Sequence 3, Appl1
6	4677.5	88.6	1002	US-09-812-633-3	Sequence 3, Appl1
7	489	9.3	330	US-09-864-761-42953	Sequence 42953, A
8	441	8.4	336	US-09-864-761-43157	Sequence 43157, A
9	267.5	5.1	515	US-09-864-761-37152	Sequence 37152, A
10	267.5	5.1	610	US-09-988-920-212	Sequence 212, App
11	266	5.0	577	US-09-764-864-883	Sequence 883, App
12	247	4.7	513	US-09-764-864-1061	Sequence 1061, App
13	244.5	4.6	338	US-09-764-864-948	Sequence 948, App
14	244	4.6	406	US-09-764-864-976	Sequence 976, App
15	242	4.6	439	US-09-864-761-33476	Sequence 33476, A
16	238.5	4.5	492	US-09-864-761-42897	Sequence 42897, A
17	237	4.5	375	US-09-864-761-35336	Sequence 35336, A
18	234.5	4.4	311	US-09-864-761-34333	Sequence 34333, A
19	234.5	4.4	340	US-09-864-761-35894	Sequence 35894, A

20	233.5	4.4	1125	US-09-974-298-114	Sequence 114, App
21	233	4.4	551	US-09-864-761-37065	Sequence 37065, A
22	230	4.4	558	US-09-864-761-47724	Sequence 47724, A
23	229.5	4.3	2783	US-09-816-669A-14	Sequence 14, Appl1
24	228	4.3	708	US-09-764-864-954	Sequence 954, App
25	227.5	4.3	323	US-09-764-864-947	Sequence 947, App
26	227.5	4.3	323	US-09-764-864-1095	Sequence 1095, App
27	226.5	4.3	361	US-09-764-864-893	Sequence 893, App
28	225.5	4.3	510	US-09-866-562-56	Sequence 56, App
29	224.5	4.3	322	US-09-864-761-36708	Sequence 36708, A
30	224	4.2	525	US-09-864-761-33633	Sequence 33633, A
31	224	4.2	536	US-09-864-761-36148	Sequence 36148, A
32	224	4.2	537	US-10-037-667-1	Sequence 1, Appl1
33	222.5	4.2	506	US-09-083-290-2	Sequence 2, Appl1
34	221	4.2	252	US-09-816-669A-12	Sequence 12, Appl1
35	220.5	4.2	383	US-09-764-864-1001	Sequence 1001, App
36	220.5	4.2	516	US-09-083-290-4	Sequence 956, Appl1
37	219.5	4.2	378	US-09-764-864-956	Sequence 957, App
38	219.5	4.2	580	US-09-925-301-1357	Sequence 1357, App
39	218	4.1	600	US-09-764-864-957	Sequence 957, App
40	217.5	4.1	516	US-09-974-298-118	Sequence 118, App
41	216.5	4.1	2005	US-09-735-367B-3	Sequence 3, Appl1
42	216.5	4.1	2053	US-09-735-367B-2	Sequence 2, Appl1
43	215.5	4.1	449	US-09-929-315-4	Sequence 4, Appl1
44	215.5	4.1	449	US-09-933-215-4	Sequence 4, Appl1
45	214.5	4.1	312	US-09-764-864-1059	Sequence 1059, App
46	214.5	4.1	329	US-09-764-864-1482	Sequence 1482, App
47	213.5	4.0	983	US-09-864-761-38000	Sequence 38000, A
48	212.5	4.0	316	US-09-764-864-1036	Sequence 1036, App
49	211	4.0	448	US-09-764-864-909	Sequence 909, App
50	211	4.0	449	US-09-929-315-6	Sequence 6, Appl1
51	211	4.0	449	US-09-933-215-6	Sequence 6, Appl1
52	211	4.0	515	US-10-037-667-4	Sequence 4, Appl1
53	210.5	4.0	468	US-09-864-761-37991	Sequence 37991, A
54	209.5	4.0	432	US-09-755-830-39	Sequence 39, Appl1
55	208	3.9	284	US-09-764-864-932	Sequence 932, App
56	207.5	3.9	428	US-09-734-329-2	Sequence 2, Appl1
57	207	3.9	359	US-09-764-864-996	Sequence 996, App
58	206.5	3.9	461	US-09-755-830-38	Sequence 38, Appl1
59	205	3.9	957	US-09-864-761-43107	Sequence 43107, A
60	205	3.9	957	US-10-025-380-1065	Sequence 1065, App
61	205	3.9	957	US-09-922-217-1065	Sequence 1065, App
62	205	3.9	957	US-09-833-263-1065	Sequence 1065, App
63	204	3.9	944	US-09-964-238-2	Sequence 2, Appl1
64	203.5	3.9	281	US-09-764-864-1076	Sequence 1076, App
65	203	3.8	269	US-09-764-864-887	Sequence 887, App
66	203	3.8	543	US-09-736-457-337	Sequence 337, App
67	203	3.8	543	US-09-902-941-337	Sequence 337, App
68	203	3.8	543	US-09-849-626-337	Sequence 337, App
69	202.5	3.8	498	US-10-037-667-5	Sequence 5, Appl1
70	202	3.8	585	US-09-764-864-1063	Sequence 1063, App
71	202	3.8	4019	US-09-854-133-425	Sequence 425, App
72	202	3.8	4019	US-09-738-973-425	Sequence 425, App
73	201.5	3.8	261	US-09-764-864-961	Sequence 961, App
74	201	3.8	294	US-09-764-864-1515	Sequence 1515, App
75	201	3.8	1367	US-09-801-368-108	Sequence 108, App
76	200	3.8	273	US-09-764-864-1421	Sequence 1421, App
77	200	3.8	571	US-09-764-864-1028	Sequence 1028, App
78	199	3.8	547	US-09-764-864-1065	Sequence 1065, App
79	198.5	3.8	277	US-09-864-761-36043	Sequence 36043, A
80	198.5	3.8	532	US-10-037-667-2	Sequence 2, Appl1
81	198.5	3.8	711	US-09-976-165-10	Sequence 10, Appl1
82	198.5	3.8	711	US-09-828-648-12	Sequence 2, Appl1
83	197	3.7	249	US-09-925-380-1179	Sequence 1179, App
84	196	3.7	488	US-09-840-787-11	Sequence 17, Appl1
85	194	3.7	605	US-09-864-761-36132	Sequence 36132, A
86	194	3.7	605	US-09-864-761-37943	Sequence 37943, A
87	193.5	3.7	241	US-09-764-864-935	Sequence 935, App
88	193.5	3.7	1464	US-09-842-777-10	Sequence 10, Appl1
89	193	3.7	322	US-09-764-864-914	Sequence 914, App
90	192.5	3.6	518	US-09-755-830-40	Sequence 40, Appl1
91	192	3.6	1325	US-09-864-761-35612	Sequence 35612, A
92	191.5	3.6	1744	US-10-108-605-25	Sequence 25, Appl1

Sequence 114, App	Sequence 37065, A	Sequence 47724, A	Sequence 14, Appl1	Sequence 954, App	Sequence 947, App	Sequence 1095, App	Sequence 893, App	Sequence 56, App	Sequence 36708, A	Sequence 33633, A	Sequence 36148, A	Sequence 1, Appl1	Sequence 2, Appl1	Sequence 12, Appl1	Sequence 1001, App	Sequence 4, Appl1	Sequence 956, Appl1	Sequence 957, App	Sequence 1357, App	Sequence 957, App	Sequence 118, App	Sequence 3, Appl1	Sequence 2, Appl1	Sequence 4, Appl1	Sequence 1059, App	Sequence 1482, App	Sequence 38000, A	Sequence 1036, App	Sequence 909, App	Sequence 6, Appl1	Sequence 6, Appl1	Sequence 4, Appl1	Sequence 37991, A	Sequence 39, Appl1	Sequence 932, App	Sequence 2, Appl1	Sequence 966, App	Sequence 38, Appl1	Sequence 43107, A	Sequence 1065, App	Sequence 1065, App	Sequence 1065, App	Sequence 2, Appl1	Sequence 1076, App	Sequence 887, App	Sequence 337, App	Sequence 337, App	Sequence 337, App	Sequence 5, Appl1	Sequence 1063, App	Sequence 425, App	Sequence 425, App	Sequence 961, App	Sequence 1515, App	Sequence 108, App	Sequence 1421, App	Sequence 1028, App	Sequence 1065, App	Sequence 36043, A	Sequence 2, Appl1	Sequence 10, Appl1	Sequence 2, Appl1	Sequence 1179, App	Sequence 17, Appl1	Sequence 36132, A	Sequence 37943, A	Sequence 935, App	Sequence 10, Appl1	Sequence 914, App	Sequence 40, Appl1	Sequence 35612, A	Sequence 25, Appl1
-------------------	-------------------	-------------------	--------------------	-------------------	-------------------	--------------------	-------------------	------------------	-------------------	-------------------	-------------------	-------------------	-------------------	--------------------	--------------------	-------------------	---------------------	-------------------	--------------------	-------------------	-------------------	-------------------	-------------------	-------------------	--------------------	--------------------	-------------------	--------------------	-------------------	-------------------	-------------------	-------------------	-------------------	--------------------	-------------------	-------------------	-------------------	--------------------	-------------------	--------------------	--------------------	--------------------	-------------------	--------------------	-------------------	-------------------	-------------------	-------------------	-------------------	--------------------	-------------------	-------------------	-------------------	--------------------	-------------------	--------------------	--------------------	--------------------	-------------------	-------------------	--------------------	-------------------	--------------------	--------------------	-------------------	-------------------	-------------------	--------------------	-------------------	--------------------	-------------------	--------------------

93	191	3.6	723	12	US-10-044-205A-32	Sequence 32, Appl
94	191	3.6	796	12	US-10-044-205A-31	Sequence 31, Appl
95	186.5	3.5	483	10	US-09-764-864-987	Sequence 987, App
96	185.5	3.5	242	10	US-09-764-864-867	Sequence 867, App
97	185.5	3.5	483	10	US-09-764-864-940	Sequence 940, App
98	184.5	3.5	333	10	US-09-764-864-958	Sequence 958, App
99	184	3.5	196	9	US-09-942-087A-30	Sequence 30, Appl
100	184	3.5	196	10	US-09-779-233-18	Sequence 18, Appl

ALIGNMENTS

RESULT 1

US-09-988-117-1
; Sequence 1, Application US/09988117
; Patent No. US20020156039A1
; GENERAL INFORMATION:
; APPLICANT: Benjamin, Thomas L.
; APPLICANT: Li, Dawei
; APPLICANT: Mok, Samuel C.
; APPLICANT: Cramer, Daniel W.
; APPLICANT: Ma, Yupo
; TITLE OF INVENTION: Diagnosing and Treating Cancer Cells
; FILE REFERENCE: 00742/066002
; CURRENT APPLICATION NUMBER: US/09/988,117
; PRIOR FILING DATE: 2001-11-16
; PRIOR FILING DATE: 2001-03-19
; PRIOR APPLICATION NUMBER: US 60/216,723
; PRIOR FILING DATE: 2000-07-07
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 1
; LENGTH: 1005
; TYPE: PRT
; ORGANISM: Homo Sapiens
US-09-988-117-1

Query Match 100.0%; Score 5277; DB 9; Length 1005;
Best Local Similarity 100.0%; Pred. No. 3.4e-251;
Matches 1005; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 MAHESERSRLGVAGEPALGDASEEDHPQVCAKCAQFTDPTFEFLAHQACSTDPPV 60
1 MAHESERSRLGVAGEPALGDASEEDHPQVCAKCAQFTDPTFEFLAHQACSTDPPV 60
61 MYIIIGQENPNNSASSSEPRBGHNNPQVMDTEHSNPPDSSSVPTDPTWGPERRGESS 120
61 MYIIIGQENPNNSASSSEPRBGHNNPQVMDTEHSNPPDSSSVPTDPTWGPERRGESS 120
121 GHFLVAATGTAAAGGGGILASPKIGATPLPESTPPAPPPPPPPPPVGGSHLNTPLI 180
121 GHFLVAATGTAAAGGGGILASPKIGATPLPESTPPAPPPPPPPPPVGGSHLNTPLI 180
181 LEEIRLVQOQROIHOMQMTBQICROYULLGSLQYVGAAPSSELPGTGTAASSKPLPLF 240
181 LEEIRLVQOQROIHOMQMTBQICROYULLGSLQYVGAAPSSELPGTGTAASSKPLPLF 240
241 SPIKPVOTSKTILASSSSSSSSGAETPKQAFHLYHPGSOHPFAGGGRSHKPTTAP 300
241 SPIKPVOTSKTILASSSSSSSSGAETPKQAFHLYHPGSOHPFAGGGRSHKPTTAP 300
301 SPALPGSTDOIASPHLAFPTTGTGLAAQCLGAARGLAETASPGDLKPKNGSELSTGEV 360
301 SPALPGSTDOIASPHLAFPTTGTGLAAQCLGAARGLAETASPGDLKPKNGSELSTGEV 360
361 MGPLEKRGGRHKCFKCAVFGSDSLQIHLNHSHTGEPKYKCNVCGNRTTGRNLKYVFNH 420
361 MGPLEKRGGRHKCFKCAVFGSDSLQIHLNHSHTGEPKYKCNVCGNRTTGRNLKYVFNH 420
421 HREKYPHVQMNPHVPEHLDVITSSGLPYGMSVDPPEKAEAEATPGGVERPLVASTT 480

|||||
421 HREKYPHVQMNPHVPEHLDVITSSGLPYGMSVDPPEKAEAEATPGGVERPLVASTT 480
481 ALSATPSTLLSTASCTATAPGLPAFNKFTVLMKAVEPKNKADENTPPGSESAISGVES 540
481 ALSATPSTLLSTASCTATAPGLPAFNKFTVLMKAVEPKNKADENTPPGSESAISGVES 540
541 STATTMLQSLKMTSLPSSWALLTNHFKSTGSPPLLCARALCASPSETSCKLOLVKIDRO 600
541 STATTMLQSLKMTSLPSSWALLTNHFKSTGSPPLLCARALCASPSETSCKLOLVKIDRO 600
601 GAVAVTSAASGAPTTAPAPSSASSSPNOCVICTRLVSCRALRLHVGONGGEPFPC 660
601 GAVAVTSAASGAPTTAPAPSSASSSPNOCVICTRLVSCRALRLHVGONGGEPFPC 660
661 VCGRAFSTRGNLRAHFVGHKASPARAQNCPICQKFTNAVTLOQHVHMLGGQIPNGG 720
661 VCGRAFSTRGNLRAHFVGHKASPARAQNCPICQKFTNAVTLOQHVHMLGGQIPNGG 720
721 TALPEGGAAQENGSEQSTVSGAGSPFOQSOQSPREELSEEEEDDEEDVTDSDS 780
721 TALPEGGAAQENGSEQSTVSGAGSPFOQSOQSPREELSEEEEDDEEDVTDSDS 780
781 LAGRSESGEKAISVRGDESEASGAEEVGTVAATAAGKEMDSNEKTTQOSSLPPPP 840
781 LAGRSESGEKAISVRGDESEASGAEEVGTVAATAAGKEMDSNEKTTQOSSLPPPP 840
841 PPSLDOPQMEQSSGVLGKKEGKPERSSSPASALTPEBEATSVTLVEELSLDEAMRK 900
841 PPSLDOPQMEQSSGVLGKKEGKPERSSSPASALTPEBEATSVTLVEELSLDEAMRK 900
901 EPGESSRRACEVCGAFPSQALAEHOKTHPKRGPLFTCVCFRCOGFLERATLKKHMLA 960
901 EPGESSRRACEVCGAFPSQALAEHOKTHPKRGPLFTCVCFRCOGFLERATLKKHMLA 960
961 HHQVOPFAPHGPONIALSLVPGCSPTSTGSLSPFRKDDPTP 1005
961 HHQVOPFAPHGPONIALSLVPGCSPTSTGSLSPFRKDDPTP 1005

RESULT 2

US-09-812-471-1
; Sequence 1, Application US/09812471
; Patent No. US20020018765A1
; GENERAL INFORMATION:
; APPLICANT: Benjamin, Thomas L.
; TITLE OF INVENTION: Diagnosing and Treating Cancer Cells
; FILE REFERENCE: 00742/062002
; CURRENT APPLICATION NUMBER: US/09/812,471
; PRIOR FILING DATE: 2001-03-19
; PRIOR APPLICATION NUMBER: US 60/216,723
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 1
; LENGTH: 1005
; TYPE: PRT
; ORGANISM: Homo Sapiens
US-09-812-471-1

Query Match 100.0%; Score 5277; DB 10; Length 1005;
Best Local Similarity 100.0%; Pred. No. 3.4e-251;
Matches 1005; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 MAHESERSRLGVAGEPALGDASEEDHPQVCAKCAQFTDPTFEFLAHQACSTDPPV 60
1 MAHESERSRLGVAGEPALGDASEEDHPQVCAKCAQFTDPTFEFLAHQACSTDPPV 60
61 MYIIIGQENPNNSASSSEPRBGHNNPQVMDTEHSNPPDSSSVPTDPTWGPERRGESS 120
61 MYIIIGQENPNNSASSSEPRBGHNNPQVMDTEHSNPPDSSSVPTDPTWGPERRGESS 120

QY	121	GHFLVAATGTAAGGGGGLLSPKLGATPLPRESTPAAPPPPPPPPPGCGSHLNPILI	180
Db	121	GHFLVAATGTAAGGGGGLLSPKLGATPLPRESTPAAPPPPPPPPPGCGSHLNPILI	180
QY	181	LEELRYLQOORQHOMQMEQICROYVLLLSIGOTVGAPAPSELPGGTASSTKPLPLF	240
Db	181	LEELRYLQOORQHOMQMEQICROYVLLLSIGOTVGAPAPSELPGGTASSTKPLPLF	240
QY	241	SPKRVQTSKTLASSSSSSSSSGAETPKQAFPHLYLHPILSQHFFSAGVGSRSHKPTAP	300
Db	241	SPKRVQTSKTLASSSSSSSSSGAETPKQAFPHLYLHPILSQHFFSAGVGSRSHKPTAP	300
QY	301	SPALPGSTDLIASPHLAPSTTGLLAOCLGAARGLEATASPELLPKPKNSGELSYEV	360
Db	301	SPALPGSTDLIASPHLAPSTTGLLAOCLGAARGLEATASPELLPKPKNSGELSYEV	360
QY	361	MGPLKPGGRHKCRFCAKYFGSDSALQILHRSHTGEBRYKCMWGNREPTTGNLKVHHR	420
Db	361	MGPLKPGGRHKCRFCAKYFGSDSALQILHRSHTGEBRYKCMWGNREPTTGNLKVHHR	420
QY	421	HREKRYPHOMNPHRYEHLDYVTTSSGLPLRGMSYPRPEKAEBAATPGGVERKPLVAST	480
Db	421	HREKRYPHOMNPHRYEHLDYVTTSSGLPLRGMSYPRPEKAEBAATPGGVERKPLVAST	480
QY	481	ALSAVESTLULSTASCTATAPGAPFAFKFYLIMKRYEKKNDENTPPSEGSAISVAES	540
Db	481	ALSAVESTLULSTASCTATAPGAPFAFKFYLIMKRYEKKNDENTPPSEGSAISVAES	540
QY	541	STATLMQLSKLMTSLPSSWMLTNHFNSTGSPFLPCARALCASPESTKLOOLVEXIDRQ	600
Db	541	STATLMQLSKLMTSLPSSWMLTNHFNSTGSPFLPCARALCASPESTKLOOLVEXIDRQ	600
QY	601	GAVAVYTAASGATTSAPAPSSSSASGPNOCVICTLRYLSCPRALRLHYGOHGERPFCK	660
Db	601	GAVAVYTAASGATTSAPAPSSSSASGPNOCVICTLRYLSCPRALRLHYGOHGERPFCK	660
QY	661	YCGRASTRGNLRAHFVGHKASPARAARONSCPICQKKTNAVTLQOHVHMLGQIPNG	720
Db	661	YCGRASTRGNLRAHFVGHKASPARAARONSCPICQKKTNAVTLQOHVHMLGQIPNG	720
QY	721	TALPEGGAAOENGBESTYSGAGSPQOQSOOPSPEELSEEEDEDEEEDVYDEOS	780
Db	721	TALPEGGAAOENGBESTYSGAGSPQOQSOOPSPEELSEEEDEDEEEDVYDEOS	780
QY	781	LARGSESGEKAISVAGDSEASGAEEVTVAAATAGKEMDSNEKTTQOSSLPPPP	840
Db	781	LARGSESGEKAISVAGDSEASGAEEVTVAAATAGKEMDSNEKTTQOSSLPPPP	840
QY	841	PSLDOPQPMBOGSSVILGCKHEBCKPERRSSPALSALTPEGATSTVIEELSLQEMRK	900
Db	841	PSLDOPQPMBOGSSVILGCKHEBCKPERRSSPALSALTPEGATSTVIEELSLQEMRK	900
QY	901	EPSESSRKACEVCGOAFPSQALAEHOKTHPKESPLFTVCRCROGFLERATLKKHMLA	960
Db	901	EPSESSRKACEVCGOAFPSQALAEHOKTHPKESPLFTVCRCROGFLERATLKKHMLA	960
QY	961	HHOVPAPHPGPNIAIALSLVPGCSPTITSGLSFPFRKDDPTIP 1005	
Db	961	HHOVPAPHPGPNIAIALSLVPGCSPTITSGLSFPFRKDDPTIP 1005	
RESULT 3			
US-09-812-633-1			
: Sequence 1, ApplicationId US/09812633			
: Patent No. US202014796A1			
: GENERAL INFORMATION:			
: APPLICANT: Benjamin, Thomas L.			
: APPLICANT: Li, David			
: APPLICANT: Mok, Samuel C.			
: APPLICANT: Cramer, Daniel W.			
: APPLICANT: Me, Yupo			
: TITLE OF INVENTION: Diagnosing and Treating Cancer Cells			
: TITLE OF INVENTION: Using Sal2			

[illegible]

OY 841 PDSLDPOPMEOGSSGVLGKKEGKPERSSPASALPEEGATSVTLVEELSLQEMARK 900
 DB PDSLDPOPMEOGSSGVLGKKEGKPERSSPASALPEEGATSVTLVEELSLQEMARK 900
 OY 901 EPESSSRKACEVCGOAFPSOALEEHOKTHPKBPLTCVCRQGFLEATLKKHMLA 960
 DB 901 EPESSSRKACEVCGOAFPSOALEEHOKTHPKBPLTCVCRQGFLEATLKKHMLA 960
 OY 961 HHQVOPFAPHGPNQIALSLVPGCSPTSTGSLSPFRKDDPTIP 1005
 DB 961 HHQVOPFAPHGPNQIALSLVPGCSPTSTGSLSPFRKDDPTIP 1005

RESULT 4
 US-09-988-117-3
 ; Sequence 3, Application US/09988117
 ; Patent No. US20020156039A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Benjamin, Thomas L.
 ; APPLICANT: Li, Dawei
 ; APPLICANT: Mok, Samuel C.
 ; APPLICANT: Cramer, Daniel W.
 ; APPLICANT: Ma, Yupo
 ; TITLE OF INVENTION: Diagnosing and Treating Cancer Cells
 ; FILE REFERENCE: 00742/066002
 ; CURRENT APPLICATION NUMBER: US/09/988,117
 ; CURRENT FILING DATE: 2001-11-16
 ; PRIOR APPLICATION NUMBER: US 09/812,633
 ; PRIOR FILING DATE: 2001-03-19
 ; PRIOR APPLICATION NUMBER: US 60/216,723
 ; PRIOR FILING DATE: 2000-07-07
 ; NUMBER OF SEQ ID NOS: 21
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 3
 ; LENGTH: 1002
 ; TYPE: PRT
 ; ORGANISM: Mus musculus
 ; US-09-988-117-3

Query Match 88.6%; Score 4677.5; DB 9; Length 1002;
 Best Local Similarity 89.0%; Pred. No. 7e-222;
 Matches 894; Conservative 25; Mismatches 83; Indels 3; Gaps 3;
 OY 1 MAHESRSSRLGVPAEPAELGDASEEDHPVCAKCAQFTDPTFEFLAHQNCSTDPV 60
 DB 1 MAHESRSSRLGVPAEPAELGDASEEDHPVCAKCAQFTDPTFEFLAHQNCSTDPV 60
 OY 61 MVTIGGQENPNSSASAPREBHSRQVMDTEHSNPPDSSGSPDPPTWGPERRGESS 120
 DB 61 MVTIGGQENPNSSASAPREBHSRQVMDTEHSNPPDSSGSPDPPTWGPERRGESS 120
 OY 121 GHFLVAATGTAAAGGGGLILASPKLGATPLPRESTPAPPPPPPPVGVSGHLNPLI 180
 DB 121 GHFLVAATGTAAAGGGGLILASPKLGATPLPRESTPAPPPPPPPVGVSGHLNPLI 180
 OY 121 GQFLVAATGTAAAGGGGLILASPKLGATPLPRESTPAPPPPPPPVGVSGHLNPLI 180
 DB 121 GQFLVAATGTAAAGGGGLILASPKLGATPLPRESTPAPPPPPPPVGVSGHLNPLI 180
 OY 181 LEELRVIOQROIHQOMQEQICROYLLGLSQIYGAAPSSELRTGTASRKRLLPLF 240
 DB 181 LEELRVIOQROIHQOMQEQICROYLLGLSQIYGAAPSSELRTGTASRKRLLPLF 240
 OY 181 LEELRVIOQROIHQOMQEQICROYLLGLSQIYGAAPSSELRTGTASRKRLLPLF 240
 DB 181 LEELRVIOQROIHQOMQEQICROYLLGLSQIYGAAPSSELRTGTASRKRLLPLF 240
 OY 241 SPIKVOATSKLASSSSSSSSGAETPKQAFHLIYHPLGSOHPFAGAGVGRSHKPTAP 300
 DB 241 SPIKVOATSKLASSSSSSSSGAETPKQAFHLIYHPLGSOHPFAGAGVGRSHKPTAP 300
 OY 241 SPIKPAQTKGTTA-SSSSSSSSGAPRQAFHLIYHPLGSOHPFAGAGVGRSHKPTAP 299
 DB 241 SPIKPAQTKGTTA-SSSSSSSSGAPRQAFHLIYHPLGSOHPFAGAGVGRSHKPTAP 299
 OY 301 SPALPGSTDOIILASPHLAFPTGTGLLAOCLGAARGLLEAASGLLKPKNSSGELTGEV 360
 DB 301 SPALPGSTDOIILASPHLAFPTGTGLLAOCLGAARGLLEAASGLLKPKNSSGELTGEV 360
 OY 300 SPALPGSTDOIILASPHLAFPTGTGLLAOCLGAARGLLEAASGLLKPKNSSGELTGEV 359
 DB 300 SPALPGSTDOIILASPHLAFPTGTGLLAOCLGAARGLLEAASGLLKPKNSSGELTGEV 359
 OY 361 MGPLEKRGGRKRCFCAVFGSDSALOILHSHTGERPKCANNCGNRFRTGNLKVAFHR 420
 DB 361 MGPLEKRGGRKRCFCAVFGSDSALOILHSHTGERPKCANNCGNRFRTGNLKVAFHR 420
 OY 360 ISSLEKRGGRKRCFCAVFGSDSALOILHSHTGERPKCANNCGNRFRTGNLKVAFHR 419
 DB 360 ISSLEKRGGRKRCFCAVFGSDSALOILHSHTGERPKCANNCGNRFRTGNLKVAFHR 419
 OY 421 HREKYPHVQMNHPVPEHLDVYITSSGLPYGMSVPEKAEBAATPGGVERRPLVASTT 480
 DB 421 HREKYPHVQMNHPVPEHLDVYITSSGLPYGMSVPEKAEBAATPGGVERRPLVASTT 480

DB 420 HREKYPHVQMNHPVPEHLDVYITSSGLPYGMSVPEKAEBAATPGGVERRPLVASTT 479
 OY 481 ALSATSESTLLSTACTATAPGLPAFNKFLVLMKAVEKKNNADEMTPGSEGSATISGVES 540
 DB 481 ALSATSESTLLSTACTATAPGLPAFNKFLVLMKAVEKKNNADEMTPGSEGSATISGVES 540
 OY 480 ALSATESLTLSTGTSTAAAPGLPTENKFLVLMKAVEKKNNADEMTPGSEGSATISGVES 539
 DB 480 ALSATESLTLSTGTSTAAAPGLPTENKFLVLMKAVEKKNNADEMTPGSEGSATISGVES 539
 OY 541 STATLMQSLKMTSLPSWMLLNHFRSTGSPFLPLCARALGASPSSTKLOOLVEKIDRQ 600
 DB 541 STATLMQSLKMTSLPSWMLLNHFRSTGSPFLPLCARALGASPSSTKLOOLVEKIDRQ 600
 OY 540 GSATRQSLKMTSLPSWMLLNHFRSTGSPFLPLCARALGASPSSTKLOOLVEKIDRQ 599
 DB 540 GSATRQSLKMTSLPSWMLLNHFRSTGSPFLPLCARALGASPSSTKLOOLVEKIDRQ 599
 OY 601 GAVAYTASAGAPTSAPRPSASSGPNQVCICRYLSCPRALRLHYGGHGERPFCK 660
 DB 601 GAVAYTASAGAPTSAPRPSASSGPNQVCICRYLSCPRALRLHYGGHGERPFCK 660
 OY 600 GAVAYTASAGAPTSAPRPSASSGPNQVCICRYLSCPRALRLHYGGHGERPFCK 658
 DB 600 GAVAYTASAGAPTSAPRPSASSGPNQVCICRYLSCPRALRLHYGGHGERPFCK 658
 OY 661 YCGRAFSTRGNLRAHFVGHKASPAARAQNSCPICQKFTTNAVTLQOHVRLHGLQIPNG 720
 DB 661 YCGRAFSTRGNLRAHFVGHKASPAARAQNSCPICQKFTTNAVTLQOHVRLHGLQIPNG 720
 OY 659 YCGRAFSTRGNLRAHFVGHKASPAARAQNSCPICQKFTTNAVTLQOHVRLHGLQIPNG 718
 DB 659 YCGRAFSTRGNLRAHFVGHKASPAARAQNSCPICQKFTTNAVTLQOHVRLHGLQIPNG 718
 OY 721 TALPREGGAOENGSRQTVSGASFPQOQSPPEELISEEEDDEEDVTDSDS 780
 DB 721 TALPREGGAOENGSRQTVSGASFPQOQSPPEELISEEEDDEEDVTDSDS 780
 OY 719 SALSBEGGAOENGSRQTVSGASFPQOQSPPEELISEEEDDEEDVTDSDS 777
 DB 719 SALSBEGGAOENGSRQTVSGASFPQOQSPPEELISEEEDDEEDVTDSDS 777
 OY 781 LAGRSESGEKALISVRCDEEASGAEEVGYAAATAGKEMDSNEKTTQSSLPPPP 840
 DB 781 LAGRSESGEKALISVRCDEEASGAEEVGYAAATAGKEMDSNEKTTQSSLPPPP 840
 OY 778 LAGRSESGEKALISVRCDEEASGAEEVGYAAATAGKEMDSNEKTTQSSLPPPP 837
 DB 778 LAGRSESGEKALISVRCDEEASGAEEVGYAAATAGKEMDSNEKTTQSSLPPPP 837
 OY 841 PDSLDPOPMEOGSSGVLGKKEGKPERSSPASALPEEGATSVTLVEELSLQEMARK 900
 DB 841 PDSLDPOPMEOGSSGVLGKKEGKPERSSPASALPEEGATSVTLVEELSLQEMARK 900
 OY 838 PDLSDPOPMEOGSSGVLGKKEGKPERSSPASALPEEGATSVTLVEELSLQEMARK 897
 DB 838 PDLSDPOPMEOGSSGVLGKKEGKPERSSPASALPEEGATSVTLVEELSLQEMARK 897
 OY 901 EPESSSRKACEVCGOAFPSOALEEHOKTHPKBPLTCVCRQGFLEATLKKHMLA 960
 DB 901 EPESSSRKACEVCGOAFPSOALEEHOKTHPKBPLTCVCRQGFLEATLKKHMLA 960
 OY 898 DGESSGRKACEVCGOAFPSOALEEHOKTHPKBPLTCVCRQGFLEATLKKHMLA 957
 DB 898 DGESSGRKACEVCGOAFPSOALEEHOKTHPKBPLTCVCRQGFLEATLKKHMLA 957
 OY 961 HHQVOPFAPHGPNQIALSLVPGCSPTSTGSLSPFRKDDPTIP 1005
 DB 961 HHQVOPFAPHGPNQIALSLVPGCSPTSTGSLSPFRKDDPTIP 1005
 OY 958 HHQVOPFAPHGPNQIALSLVPGCSPTSTGSLSPFRKDDPTIP 1002
 DB 958 HHQVOPFAPHGPNQIALSLVPGCSPTSTGSLSPFRKDDPTIP 1002

RESULT 5
 US-09-812-471-3
 ; Sequence 3, Application US/09812471
 ; Patent No. US20020018765A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Benjamin, Thomas L.
 ; TITLE OF INVENTION: Diagnosing and Treating Cancer Cells
 ; FILE REFERENCE: 00742/062002
 ; CURRENT APPLICATION NUMBER: US/09/812,471
 ; CURRENT FILING DATE: 2001-03-19
 ; PRIOR APPLICATION NUMBER: US 60/216,723
 ; PRIOR FILING DATE: 2000-07-07
 ; NUMBER OF SEQ ID NOS: 21
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 3
 ; LENGTH: 1002
 ; TYPE: PRT
 ; ORGANISM: Mus musculus
 ; US-09-812-471-3

Query Match 88.6%; Score 4677.5; DB 10; Length 1002;
 Best Local Similarity 89.0%; Pred. No. 7e-222;
 Matches 894; Conservative 25; Mismatches 83; Indels 3; Gaps 3;
 OY 1 MAHESRSSRLGVPAEPAELGDASEEDHPVCAKCAQFTDPTFEFLAHQNCSTDPV 60
 DB 1 MAHESRSSRLGVPAEPAELGDASEEDHPVCAKCAQFTDPTFEFLAHQNCSTDPV 60
 OY 1 MAHESRSSRLGVPAEPAELGDASEEDHPVCAKCAQFTDPTFEFLAHQNCSTDPV 60
 DB 1 MAHESRSSRLGVPAEPAELGDASEEDHPVCAKCAQFTDPTFEFLAHQNCSTDPV 60
 OY 61 MVTIGGQENPNSSASAPREBHSRQVMDTEHSNPPDSSGSPDPPTWGPERRGESS 120
 DB 61 MVTIGGQENPNSSASAPREBHSRQVMDTEHSNPPDSSGSPDPPTWGPERRGESS 120
 OY 61 MVTIGGQENPNSSASAPREBHSRQVMDTEHSNPPDSSGSPDPPTWGPERRGESS 120
 DB 61 MVTIGGQENPNSSASAPREBHSRQVMDTEHSNPPDSSGSPDPPTWGPERRGESS 120


```
QY      841 PDSLDPOPMEGSSGVLGKKEGGKPERSSSPASALTPREGATSVTLVEELSLQEMARK 900
      838 PDIIDHPQMEGTSIDVSGAMEEBAKLEBISPMALTOGEGSTFLVEELNPEAMKK 897
QY      901 EPESSSRKACVCGQAPPSQALAEHQKTRKEGLFTVFCRQGLERATLKKHMLLA 960
      898 DPSSSRKACVCGQSPPTQALAEHQKTRKEGLFTVFCRQGLERATLKKHMLLA 957
QY      961 HHQVPPAPHPQNIATLSTLVPGCSSSTSPSPRPKDDPTIP 1005
      958 HHQVPPAPHPQNIATLSTLVPGCSSSTSPSPRPKDDPTIP 1002

RESULT 7
US-09-864-761-42953
; Sequence 42953, Application US/09864761
; Patent No. US20020048763A1
GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: Aecm1ca-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
SEQ ID NO 42953
LENGTH: 330
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AC007498.2
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.1
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.2
```

```
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.2
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.4
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.4
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 0.96
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 0.94
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.97
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.5
; OTHER INFORMATION: EST_HUMAN HIT: AUI31453.1, EVALUATE 1.00e-105
; OTHER INFORMATION: SWISSPROT HIT: P39806, EVALUATE 4.00e-43
US-09-864-761-42953

Query Match          9.3%; Score 489; DB 10; Length 330;
Best Local Similarity 33.3%; Pred. No. 1.5e-17;
Matches 136; Conservative 51; Mismatches 114; Indels 108; Gaps 14;

QY      529 SEGSAISGVAESSTATTLMQSLKMTSLPSWALLTNHFKSTGSEFPLICARALGASPSETS 588
      3 SSPAADCGPAGSATTFTNPL-----LP---LMSQFKA-----KFRGGLDLSAQASETS 49
DB      589 KLOOLYERKIDROGAVAVTSASGAPTTSPAPSSSSASGPNOCVICLRVYSCPALRLHY 648
      KLOOLYERIDKK-----ATDPNECIIHRVLSCOSALKMRY 85
QY      649 GONGGERPFKCKVCGRFSTGTGNLRAHFVGHKASPARAARONSPICOKFTNAVTLQOHV 708
      RTHTGERPFKCKICGRATTTGNLTKHYSVHRAMPRLRVQHSFCPICOKFTNAVYLOHI 145
DB      709 RMHLGGIIPNGTALPDEGGAOENGSEOSTVSGAGSPFOQSOQSPSEELSEEEED 768
      RMHGGIIPN--TPVPPSYSESM-----SDGGSF-----DEKNDD 180
QY      146 RMHGGIIPN--TPVPPSYSESM-----SDGGSF-----DEKNDD 180
      EEEEDVTDEDSLGRGSESGEKAISVRGDSFASGAEFVTAANAATAGKEMDNEK 828
DB      181 LDNFSDPMED--CEPGSIPTPKSADASODLSLSPDLLEMSITALENQMKTINGLA 238
      TPOQSSLPPEPPPSLDPOPMEGQS-----SGVLG---KEEGKPERSSSPAS- 875
QY      239 EQLOASL-----KSVENGSIEGDVLITNDSSVSGDMESQAGSPALISESTSM 286
      -ATTPREGATSVTLVEELSLQEMARKKEGESSSRKACVCGQAPPSQAA 923
DB      876 -ATTPREGATSVTLVEELSLQEMARKKEGESSSRKACVCGQAPPSQAA 923
      287 QALSPSNS-----TQEFHKSPISEKP-----QRAVPSEFA 317

RESULT 8
US-09-864-761-43157
; Sequence 43157, Application US/09864761
; Patent No. US20020048763A1
GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
FILE REFERENCE: Aecm1ca-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
```


PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
SEQ ID NO 43157
LENGTH: 336
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AL034420.16
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.72
OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.1
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.75
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 0.76
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 0.78
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.75
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.86
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.81
OTHER INFORMATION: EST_HUMAN HIT: AM270757.1, EVALUE 7.00e-68
OTHER INFORMATION: SWISSPROT HIT: P39806, EVALUE 8.00e-39
US-09-864-761-43157

Query Match 8.4%; Score 441; DB 10; Length 336;
Best Local Similarity 34.3%; Pred. No. 3.4e-15;
Matches 119; Conservative 41; Mismatches 89; Indels 98; Gaps 11;
QY 585 SETSKLOOIVEKIDROGAVAVTSAASGAPTTAPAPSSASSGPNOCYICLRVLSCPRAL 644
DB 62 SETKLQOQLEVIDK-----ATTDPNECLDCHRVLSGSSSL 97
QY 645 RLHVGCGGGERPKKVCGRAFSTRGNLRAHFVGHKASPARAARNSCPICQKFTNAVTL 704
DB 98 KMHYRTHTGERPFCKICGRARSTKGNLKLTHLVHRTNTSIKTHSCPIQCKFTNAVNL 157
QY 705 QQAHYMLHGGQIPNGGTALPEQSGAAQENGSEQTVSGAGSFPOQSOQSPSEBELSEEE 764
DB 158 QQHRLMHGGQIPN-TLPE- NPDFTGSEPMITVGENG-----TGAICHD 202
QY 765 EEDDEEEDVYDEDSLGRGSESGEKEAISVRGDESEASCAEEVGTV-----AAATA 819
DB 203 VIESIDVEE-----VSSQEAPESSSKVPTPLPSIHNSASPTL 238
QY 820 GKENDSNKTTQOSSLP-----PPDSLDOPQPMEOG-----SSGVLGKKEEG 865
DB 239 GFAM-----MASIDAPGKKGPAPFNLDROGSRENGVSVDGILTNNSSSLMGDE--- 287
QY 866 KPERSSSPASALTPEGEATSVTLVEELSLDEAMRKREGESSSRKACE 912
DB 288 --YGRSPDILETTSPQALSPANSQAESINS--KSPDAGSKAESSE 329

RESULT 9
US-09-864-761-37152

Sequence 37152, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: A60m1ca-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
SEQ ID NO 37152
LENGTH: 515
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AL049646.17
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.6
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.9
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.8
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 2.3
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.7
OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 2.6
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.99
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.9
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.8
OTHER INFORMATION: EST_HUMAN HIT: AW953464.1, EVALUE 0.00e+00
OTHER INFORMATION: SWISSPROT HIT: P52736, EVALUE 0.00e+00
OTHER INFORMATION: EST_HUMAN HIT: AW953464.1, EVALUE 0.00e+00
US-09-864-761-37152

Query Match 5.1%; Score 267.5; DB 10; Length 515;
Best Local Similarity 20.9%; Pred. No. 1.6e-06;


```
Matches 160; Conservative 60; Mismatches 232; Indels 313; Gaps 30;
OY 237 LPLPSPIKPVQTSKLTISSSSSSSSSSSGAETPKQAFHLHPGSGHPFSAAGVGRSHKP 296
Db 13 MFLGRTK---KRTLGAFS-----RPPQR-----QPVSSRN-----GLGVTELEA 49
OY 297 TPAPSPALPGSTDLI-----ASPHLAPSTTGLLAACGLGAARGLEATVASP 343
Db 50 SPQAS-GNPETDRLKRIEVLGFTVNCGEGLSTSKMTNLISHQ----- 94
OY 344 GLLKPKNGSELSTGEYWG-----PLEKPGGRHK-----CRFCAKVFQSDALQ 387
Db 95 -----RIHSGEKPY--VGVYCEKGFSLKSLARHOKAHSGEPIVCEGGRGNKSTLI 147
OY 388 IHLRSHTGERPRKCNVCGNRFTRGNLKVHFRHREKYPHVQNNPHVPEHLDYITSSG 447
Db 148 IHERTHSGEPRYMCSECGRGFSOKSNLTIHORTHSGEKP-----YVCRECG 193
OY 448 LPYGMSVPRPEKAEDEAATPGGVERKPLVASTATLATESLTLLSTAGATAPGLPAFN 507
Db 194 KGFSSKSAVVRHQET-----HLEKRTIVCSDCGLGFSDRSNLIS----- 232
OY 508 KPYLMKAVEKKNADENTPPSEGSATIS--GVAESSTATLMOLSKLMTSLPWSALLTNHF 565
Db 233 -----HQRTHSGEKPYACKECGRCFRQRTTLV-----NH- 261
OY 566 KSTGSPFLPLCARALGASPESTSKLOQLVEKIDROGAVALVTSAGAPTTAPAPSSSAS 625
Db 262 QHTHSKEKPYVCGVCGHSFQNSNLT-----ISHR-----RTHTG 295
OY 626 SGPNOVCITLRYLSCPRALRLHYGOHGERPKCKVCGRAFSTRGNLRAHFVGHKASPA 685
Db 296 EKPYVCGVGRGFSLSKSHLNHONIHSGEKPIVCKDCGRGFSQSNL-----IHRTHSG 351
OY 686 RAONSPICOKKFTNATLQOHVBMHGGQIPNGSTALPEGGAOENSGEOSTVSGAGS 745
Db 352 EKPYVCGVGRGFSQSNLVAHORTHSG-----ERPYVCEG- 389
OY 746 FPQOOSQPSPEELSEDEEDEDVTDDESLAGSGESGEEKATSVGDSEASG 805
Db 390 -----RQFSHQ----- 395
OY 806 ABEVGTVAATAAGKEMDSNEKTTQOSSLP PPPPSLIDOPQPMEGSSGVLGKEEG 865
Db 396 -----AGLRHRRKHREK-----PYMCRQCGLG-FGNK----- 423
OY 866 KPERSSPASALTPPEGATVTLVEELSLQAMRKPEGSSSRKACEVCGOAFPSQALE 925
Db 424 -----SALTTHKRAHS-----BEKP-----CYCRECGQGLQKSHLT 455
OY 926 EHOKTHPKEGPLFTCVFCRQGLERATLKKH--MLAHNQ--VOP 966
Db 456 LHOVHTHGEKP--YVCKTCGRGFSLSKSHLSRHKRTTSVHHRLPVOP 499
RESULT 10
US-09-989-920-212
; Sequence 212, Application US/09989920
; Patent No. US20020172957A1
; GENERAL INFORMATION:
; APPLICANT: Macdon, Roberto
; APPLICANT: Recipon, Herve
; APPLICANT: Chen, Sei-Yu
; APPLICANT: Sun, Yongming
; APPLICANT: Liu, Chenghua
; TITLE OF INVENTION: Compositions and Methods Relating to Lung Specific Genes and Prot
; FILE REFERENCE: DEX-0291
; CURRENT APPLICATION NUMBER: US/09/989,920
; CURRENT FILING DATE: 2001-11-21
; PRIOR APPLICATION NUMBER: 60/252,500
; PRIOR FILING DATE: 2000-11-22
; NUMBER OF SEQ ID NOS: 284
; SOFTWARE: PatentIn version 3.1
```

```
; SEQ ID NO 212
; LENGTH: 610
; TYPE: PRT
; ORGANISM: Homo sapien
US-09-989-920-212
Query Match 5.1%; Score 267.5; DB 9; Length 610;
Best Local Similarity 20.4%; Pred. No. 1.9e-06;
Matches 129; Conservative 74; Mismatches 187; Indels 241; Gaps 23;
OY 365 EKPGRRHKRCFCAKVFQSDALQIHLRSHTGERPRKCNVCGNRFTRGNLKVHFRHREK 424
Db 132 ERP--YVCDVCGKTFRRNAGLKVHRLHTGKERPYCDVCGKAYISNLSKNHKGTHLGE 188
OY 425 YP-----HVQNNPHVPEHLDYITSSGIPYGMV-----PPEK 458
Db 189 KPYKCYCEKSFNYSALQNHKRIHTREKPFQDECGKAFRNNSGLKVHRIHTGERPK 246
OY 459 AEDEAATPGGVEKRPPLVASTATLATESLTLLSTAGATAPGLPAFNKAVLEPK 518
Db 249 CEE-----CGKAYISLSLHNKSVHGERPF----- 275
OY 519 NKADENTPPSEGSATISGVAESSTATLMOLSKLMTSLPWSALLTNHF--STGSPFLPL 575
Db 276 -KCDE-----CEKAFIT-----YRITLHNKRVHLGEKPYKCDV 307
OY 576 CARALGASPESTSKLOQLVEKIDROGAVALVTSAGAPTTAPAPSSSASGPNQCVICL 635
Db 308 CEKSF--VYTSLLSOHRRVHTRE-----KPYECDRCE 337
OY 636 RVLSCPRALRLHYGOHGERPKCKVCGRAFSTRGNLRAHFVGHKASPARAONSPICQ 695
Db 338 KVRNNSSLKVKHRIHTGERPYECDVCGKATISHSL-----INHKSTHCKTTHTCDECG 393
OY 696 KFTNATLQOHVBMHGGQIPNGSTALPEGGAOENSGEOSTVSGAGSFPQOOSQOPS 755
Db 394 KAFSSRTLLISHKRVHIG-----EKPFKVECGKSFSSYS- 428
OY 756 PEEELSEDEEDEDVTDDESLAGSGESGEEKATSVGDSEASGAEEVGTVA 815
Db 429 ----LSQHRRIHTGERPYVCDRCGKAFRNS-----S 456
OY 816 AATAGKEMDSNEKTTQOSSLP PPPPSLIDOPQPMEGSSGVLGKEEGKPERSSPAS 875
Db 457 GLTVHRIHTGER-----PYE-----CDECGKATISHS--- 484
OY 876 ALTPPEGATVTLVEELSLQAMRKPEGSSSRKACEVCGOAFPSQALEEHOKTHPKEG 935
Db 485 -----SLINHSVHOG--KOPYN-----CE-CGKSFNYSVLDOHRIHTGKK 524
OY 936 PLFTCVFCRQGLERATLKKHMLAHNQVOP 966
Db 525 P-YRCNDC-AHTPNAT--ADLMKVDBEEP 550
RESULT 11
US-09-764-864-883
; Sequence 883, Application US/09764864
; Patent No. US20020132753A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT223
; CURRENT APPLICATION NUMBER: US/09/764,864
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1792
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 883
; LENGTH: 577
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-764-864-883
```


Query Match	4.6%	Score 244.5;	DB 10;	Length 338;
Best Local Similarity	23.8%	Pred 1.4e-05;		
Matches	85;	Conservative 34;	Mismatches 145;	Indels 97;
				Gaps 10

RESULT 14
US-09-764-864-976
Sequence 976, Application US/09764864
Patent No. US20020132753A1

ORGANISM: HOTT
JS-09-764-864-976

365 EKRGGRHKRCRECAKVFSGDSALQIHLRSTHTGERPYCNVCGNFTTGRNKLVFHFRHREK 424
 98 EKP---YECPECGKAFASEKSRLLKHQRTHTGEEKPYKCDGCDKAFAKSGIRIHQRTHTG 154

RESULT 15
US-09-864-761-33476
; Sequence 33476, Application US/09864761
; Patent No. US20020048763A1

; PRIOR APPLICATION NUMBER: PCT/US01/0066/
; PRIOR FILING DATE: 2001-01-30


```

? PRIOR FILING DATE: 2000-09-27
? PRIOR APPLICATION NUMBER: PCT/US01/00666
? PRIOR FILING DATE: 2001-01-30
? PRIOR APPLICATION NUMBER: PCT/US01/00667
? PRIOR FILING DATE: 2001-01-30
? PRIOR APPLICATION NUMBER: PCT/US01/00664
? PRIOR FILING DATE: 2001-01-30
? PRIOR APPLICATION NUMBER: PCT/US01/00669
? PRIOR FILING DATE: 2001-01-30
? PRIOR APPLICATION NUMBER: PCT/US01/00665
? PRIOR FILING DATE: 2001-01-30
? PRIOR APPLICATION NUMBER: PCT/US01/00668
? PRIOR FILING DATE: 2001-01-30
? PRIOR APPLICATION NUMBER: PCT/US01/00663
? PRIOR FILING DATE: 2001-01-30
? PRIOR APPLICATION NUMBER: PCT/US01/00662
? PRIOR FILING DATE: 2001-01-30
? PRIOR APPLICATION NUMBER: PCT/US01/00661
? PRIOR FILING DATE: 2001-01-30
? PRIOR APPLICATION NUMBER: PCT/US01/00670
? PRIOR FILING DATE: 2001-01-30
? PRIOR APPLICATION NUMBER: US 60/234,687
? PRIOR FILING DATE: 2000-09-21
? PRIOR APPLICATION NUMBER: US 09/608,408
? PRIOR FILING DATE: 2000-06-30
? PRIOR APPLICATION NUMBER: US 09/774,203
? PRIOR FILING DATE: 2001-01-29
? NUMBER OF SEQ ID NOS: 49117
? SOFTWARE: Annonch Sequence Listing Engine vers. 1.1
? SEQ ID NO 36336
? LENGTH: 375
? TYPE: PRT
? ORGANISM: Homo sapiens
? FEATURE:
? OTHER INFORMATION: MAP TO AC011451.2
? OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.4
? OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.4
? OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.8
? OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.3
? OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.2
? OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.5
? OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.7
? OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.96
? OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.4
? OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.3
? OTHER INFORMATION: EST HUMAN HIT: AF150239.1, EVALUO 1.00e-102
? OTHER INFORMATION: SWISSPROT HIT: Q14585, EVALUO 1.00e-124
US-09-864-761-36336
Query Match 4.5%; Score 237; DB 10; Length 375;
Best Local Similarity 24.1%; Pred. No. 3,6e-05;
Matches 84; Conservative 43; Mismatches 170; Indels 52; Gaps 8;
372 KCFCFCAKVFESDSALQHLRSHTEGPRKCNVCNCFTRGNLKVFNHNRREKYPHQMN 431
Db 39 ECHEGCGAFAVDHSLKSHIRSHGSKPQCKCGKGAFAHFLACPKKHNKPTEE-----K 92
QY 432 PRRVPEHIDLYVITSSGLPYGMSVPEKAEAEATPGGVERKPLVASTALSAESTLTL 491
Db 93 PYECKEOTKAFSCSSFFRAHMKHNIGKTNYECKEGKG-----FSCSSSLT-- 138
QY 492 STSAGTATAGLPAFNKFLVLMKAVEPKKADENTPPGSGSAISGVAE---SSTATIMQ 547
Db 139 -----EKKRIHSGDKPYECKEGKGFSCSSSSLSKHNKRHSGDKPYE 179
QY 548 LSKLMTSLPFWALLTNFK--STGSPFLDLARALASPESTSL--QOLVEKIDRQAVA 604
Db 180 CKECGAFASSSHLITIRINTGEKPYE--CKECKAISESSKLVYHGRTHGEKPYCK 237
QY 605 VTSAGAGPTSAAPAPSSASSGPNQCVICLVLSCPRALRLHYGONGERRPRCKVCGR 664
Db 238 ECGAAYNCPSLSTIHMKHNIGKEKPYECLECGKAFVLTPTSLNTHVKMQNSREKPYECKG 297

```


Oy 665 AESTGNLAAHFVGHKASPAARQNSCPICQKFTNTAVTLQOHVPMHIG 713
Db 298 AFSCSSFRAHVRDH---TGKIQYECKEGCKTFSSRSSSLTEHLRTHSG 342

RESULT 18
US-09-864-761-34333
Sequence 34333, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Penn, Sharon G.
APPLICANT: Hanzel, David R.
APPLICANT: Hanzel, David R.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: Aecm1ca-x-1
CURRENT FILING DATE: 2001-05-23
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US/09/864,761
PRIOR FILING DATE: 2000-02-04
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263,6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
SEQ ID NO 34333
LENGTH: 311
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AC008165.2
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.5
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.9
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.97
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.2
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.7
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 2.3
OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 2.5
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.7

OTHER INFORMATION: SWISSPROT HIT: P52742, EVALUATE 1.00e-115
OTHER INFORMATION: EST_HUMAN HIT: A0133067.1, EVALUATE 3.00e-85
US-09-864-761-34333

Query Match 4.4%; Score 234.5; DB 10; Length 311;
Best Local Similarity 23.5%; Pred. No. 3.9e-05;
Matches 85; Conservative 37; Mismatches 142; Indels 97; Gaps 11;

Oy 365 EKPGGRHKCFKCAKVFSGDSALQTHLRSHHTGPRPKCNCVNGNFTTRGNLKVPHRH-RE 423
Db 27 EKP---HGCHLCGKAFTICSDLRKHERHIL-----GDKPYG----- 115
Oy 424 K-----YPHVQMNPHVPEHLDVYTSSGLPYGNSVPEKAEAEATPGGVK 473
Db 84 KAQICHLGKAFTICSDLRKHERHIL-----GDKPYG----- 115
Oy 474 PLVASTALSTESLTLLSTAGTATDGLPAPNKFVLKMAVEPKKADENTPPGSGSA 533
Db 116 -----CLLCGR-----AFSKCSYLROHERHTHNGEKPVECHLCGKA 150
Oy 534 ISGVAESSTATLMQSKMTSLPMSALLTNHKS--TGSFP--LPLCARLGAASPETSKL 590
Db 151 FSHCSH-----LROHERSHNGEKPVECHLCGKA--TESSVLKR 187
Oy 591 QQLVEKIDROGAVAVTSAASGAPTSAPAPSSASSGPNOCVTCRLVLSCPRLRLHYGQ 650
Db 188 HERHTGKEKPYECGKAFETSSDLRHERHTHNGEKPVECHLCGKAFTNHSVLRHERBT 247
Oy 651 HGERPFRCKVCGRAFTSRGNLRHAFVGHKASPAARQNSCPICQKFTNTAVTLQOHVRM 710
Db 248 HTGEKPYECNICGKAFTNSYNFRLHRRVH---TGKPYVCPLCGKAFTFNLRQHERT 303
Oy 711 H 711
Db 304 H 304

RESULT 19
US-09-864-761-35894
Sequence 35894, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Penn, Sharon G.
APPLICANT: Hanzel, David R.
APPLICANT: Hanzel, David R.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
FILE REFERENCE: Aecm1ca-x-1
CURRENT FILING DATE: 2001-05-23
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US/09/864,761
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263,6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30

Db 884 DQNDSDPTPKKKMRK-----TENGYACDLCDKIFOKSSS-----LLRHKE 926
Qy 599 RQGAIVATSAASGAPTSAAPASSASSGPNOCVLCRLVLSCRALRLHYQHGGGERPK 658
Db 927 HTG-----KRPHECGICKKAFKHKHLLHMKRLHSGEKYQ 962
Qy 659 CKYCGRAFSTRGNLRAHFVGHKASPAAR-AQNSCPICOKKFTNAVTLQOHVRLHGGQIP 717
Db 963 CDKCGKRFSGHSISQH-MNIRYCYCKREAEERDSTGEAEGLISNEHY-----1012
Qy 718 NGGTALPEGGAQENSGEOSTVSGAGSPQOQSOQPSPEELSEEE-----EEDDE 770
Db 1013 -GARASPSQGSDDR---ESTLREDEDESEKEEEDKEMEELQEKKECKPQGDDEEE 1068
Qy 771 EEDYDVDEDSLAGSGESGGERKIS---VRGSEF--AGAAEEVGVAAATAGKEDS 825
Db 1069 EEEVEEEVEEV---EAAENEGEAKTEGLMKDDRAESQASSLGQKVG-----BSSEQVS 1118
Qy 826 NEKTTQ 831
Db 1119 EKKTNE 1124

RESULT 21
US-09-864-761-37065
; Sequence 37065, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; PRIOR FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263,6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29

; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 37065
; LENGTH: 551
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC005261.1
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.8
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.5
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.3
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.6
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 2.2
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.8
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 4
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.9
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.3
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 2.3
; OTHER INFORMATION: SWISSPROT HIT: P51522, EVALUATE 0.00e+00
; OTHER INFORMATION: EST_HUMAN HIT: AF150239.1, EVALUATE 1.00e-105
; OTHER INFORMATION: EST_HUMAN HIT: AF150239.1, EVALUATE 1.00e-105
US-09-864-761-37065

Query Match 4.4%; Score 233; DB 10; Length 551;
Best Local Similarity 21.7%; Pred. No. 8.4e-05;
Matches 97; Conservative 46; Mismatches 193; Indels 110; Gaps 12;

Qy 365 EKPGGRHKRCFCAYKVGSGSALOHLRSHGGERPKYKNCVGNRTTGNLKVHFIHRH---421
Db 180 ERP---YECMECGKAFNRKSYLTQHQRHISGEKPYKCNCEGKAPTHRSNFVNLNRRHGE 236
Qy 422 -----REKYPHYOMNPH-----PVPEHLDYVI-----TSSGLPYGM 452
Db 237 KSFVCTEGCGVFRNHRPGLRHYVHSGENPYECLECKVYKRNRYLIMHOQTHG-----291
Qy 453 SVPRKAEAEATPGGVYERKPLVASTLSTATESLTLTSLAGTATAPGRLPAFKKFLYM 512
Db 292 -----EKPYECSECGKVFLESALIH---HYVITGKPFCELECG 329
Qy 513 KAVEPKN--KADENTPPGSEGSASISGVAESSTATLMLDSKMTSLPSWALLLNHKKSTGS 570
Db 330 KAFNHRSYLKRQRIHTGKRPVCS-----ECGKAFYTCSTPTILHKNRAHTGEK 378
Qy 571 PPLPLCARALGASPESTSKLOOLVEKIDROGAIVATSAASGAPTTSAAPASSASSGPNQ 630
Db 379 FECKRCGKAF-----SNRKDLIRH-----SIHTGEKPYE 408
Qy 631 CVICLRVLSCPRALRLHYQHGGGERPKCYVCGRAFSTRGNLRAHFVGHKASPAARAQNS 690
Db 409 CVECGKAFTRMGSLTRHKKRIHSGEKPYECVCEGKSCFCWSTMLIRRAIHTGKPYK----464
Qy 691 CPICOKKFTNAVTLQOHVVMHHLGGQIPNGSTALPEGGAQENSGEOSTVSGAGSPQOQ 750
Db 465 CSECGKAFTRSSSLTLOHQRNHT-GKNPISVTDVGRPFTSGOTSVTLRELLGKDFLNVTT 523
Qy 751 SQQPSPEELSEEEDEDEEEDVT 776
Db 524 EANILPEETSSASDQPYQRETPQVS 549

RESULT 22
US-09-864-761-47724
; Sequence 47724, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761


```
/ CURRENT FILING DATE: 2001-05-23
/ PRIOR APPLICATION NUMBER: US 60/180,312
/ PRIOR FILING DATE: 2000-02-04
/ PRIOR APPLICATION NUMBER: US 60/207,456
/ PRIOR FILING DATE: 2000-05-26
/ PRIOR APPLICATION NUMBER: US 09/632,366
/ PRIOR FILING DATE: 2000-08-03
/ PRIOR APPLICATION NUMBER: GB 24263,6
/ PRIOR FILING DATE: 2000-10-04
/ PRIOR APPLICATION NUMBER: US 60/236,359
/ PRIOR FILING DATE: 2000-09-27
/ PRIOR APPLICATION NUMBER: PCT/US01/00666
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00667
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00664
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00669
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00665
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00668
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00663
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00662
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00661
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00670
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: US 60/234,687
/ PRIOR FILING DATE: 2000-09-21
/ PRIOR APPLICATION NUMBER: US 09/608,408
/ PRIOR FILING DATE: 2000-06-30
/ PRIOR APPLICATION NUMBER: US 09/774,203
/ PRIOR FILING DATE: 2001-01-29
/ NUMBER OF SEQ ID NOS: 49117
/ SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
/ SEQ ID NO 47724
/ LENGTH: 558
/ TYPE: PRT
/ ORGANISM: Homo sapiens
/ FEATURE:
/ OTHER INFORMATION: MAP TO AC011451.2
/ OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.1
/ OTHER INFORMATION: SWISSPROT HIT: P51523, EVALU0 0.00e+00
/ OTHER INFORMATION: EST_HUMAN HIT: AF150239.1, EVALU0 6.00e-92
/ OTHER INFORMATION: EST_HUMAN HIT: AF150239.1, EVALU0 6.00e-92
US-09-864-761-47724

Query Match 4.4%; Score 230; DB 10; Length 558;
Best Local Similarity 19.0%; Pred. No. 0.00012;
Matches 118; Conservative 65; Mismatches 204; Indels 234; Gaps 21;

366 KGGGRKRCRCACVFGSDSLQILHRSHTGERPRKCNVCGNFFTRGNLKVHHRH-REK 424
142 KP--YKCHKCGKTFYVPSGFLFHVHRTHTGKPYCKCEKGKAFGTSAGLIEHTRCHAREK 198
425 -----YP-----HYVMNPHVP-ELHDVITSSG.PYCGSVPEPEKAEAEATP 466
199 TRKCHDCAFIYPSLFGHLRVINGEKRYEHKEK-----GKAFG-----TS 240
467 GGGVERKPLVASTTALSTATES-----LTLSTAGATATAGLPAFNKRVLMKAVEPKN 519
241 SGVIEDR---RSNTGKGRFDCCQCKVFSFSLFPAHLRTHGKEKFKCYKCK----- 291
520 KADEMTPRPSGSAISVASSSTATLMQSLKMTSLPSMALLTNPFSKSGSPPLICARA 579
292 -----PTSSACLRIHMRHTHTERTERLYQCKKCGKAFTRKCSYLTFKHLR----- 332
580 LGASPSSETKLQOLVERIKIDRGAAVAVTSAASGAPTTSA PAPSSASSGPNOCVIGLRYLS 639
```

```
Db 333 -----THAGEKPYECMCKGKAF 350
QY 640 CPRALRLHYGCHGERPRCKVCYGRAFSTRGNIRAFVGHKASBPAAQNSCPICQCKFT 699
Db 351 ERSYLTKHLRRHSGERPYECKKCGKAFTERSDTLTKHLRRTGDKPYEKD----GKAFV 406
QY 700 NAVTLQOHVRMHLGGQIINGFALPEGGAQENGSEQSVSGASGSPQQQSQSPBEE 759
Db 407 VSSSLVDHLRTHGTGK-----PYKNACEKAYRSCLYT-----QH 442
QY 760 LSEEEEEEDEEEDVDTEDSLARGSESGGKKAISVRGDSSEASGAEEYGVAAATA 819
Db 443 LKTHAAEK-----TSCNACGNSFRN----- 463
QY 820 GKENDSEKTKQSSLP PPPPPDSLDQPPMEQSSGVLGKKEGCKPERSSPSAALTP 879
Db 464 --SMCFHRLKTKTKI-----KPYCKDCGKAFTC 491
QY 880 EGEATSVTLVELSLQEARKEPGESSSRKACEVCGQAPPSOALFEHOKTHPEKPLFT 939
Db 492 HSDLTN-----HVRHTGKPYK--CKEKGARTSSGRIOHLRTHGKEK- FE 537
QY 940 CVFCRQGF---LERATLKKH 956
Db 538 CDQCKAKAFSFSARIHLKTH 558

RESULT 23
US-09-816-669A-14
/ Sequence 14, Application US/09816669A
/ Patent No. US20020137019A1
/ GENERAL INFORMATION:
/ APPLICANT: GARABEDIAN, Michael
/ APPLICANT: TANEMJA, Samir
/ APPLICANT: HITTELMAN, Adam
/ APPLICANT: MARKUS, Steven
/ TITLE OF INVENTION: METHOD FOR SCREENING TRANSCRIPTIONAL COREGULATORY PROTEINS OF
/ TITLE OF INVENTION: TRANSCRIPTION FACTORS, AND ANDROGEN RECEPTOR TRANSCRIPTIONAL
/ FILE REFERENCE: GARABEDIAN-1.1A
/ CURRENT APPLICATION NUMBER: US/09/816,669A
/ CURRENT FILING DATE: 2001-03-26
/ PRIOR APPLICATION NUMBER: 60/225,618
/ PRIOR FILING DATE: 2000-08-15
/ PRIOR APPLICATION NUMBER: 60/191,768
/ PRIOR FILING DATE: 2000-03-24
/ NUMBER OF SEQ ID NOS: 20
/ SOFTWARE: PatentIn version 3.1
/ SEQ ID NO 14
/ LENGTH: 2783
/ TYPE: PRT
/ ORGANISM: Human
US-09-816-669A-14

Query Match 4.3%; Score 229.5; DB 10; Length 2783;
Best Local Similarity 20.2%; Pred. No. 0.00067;
Matches 229; Conservative 117; Mismatches 397; Indels 391; Gaps 50;

16 GEPAEIAGD-----ASESDHPQVCAKCAQFDPTEFLIAHO-NACSTDPPTMVTIIGENP 70
Db 1396 GERRELANDNRTISNLTNQ--CKKCSLVFORITFDLIKQKLYKDEDE---EGDDDS 1449
QY 71 NN-----SSASSEPRPEGHNPNVMDTEHNSNPDSGSSVPTDPTWGPERGE 117
Db 1450 QNEDSMDAWEILPPTSSSCSPKMP-----SQAYSAFASANNMTASAFQLTAAEAE 1500
QY 118 ESSGHFLVAATGTAAAGGGGGLILASPKLGATP-LPEESTPAPPPPPPPPPGVSGHLN 176
Db 1501 E---LATENSKTEAGD-----EKPKLAEPASAPQNOTQERKQGPKP----- 1538
QY 177 IPLLBELRVLOOROIHOMQMT-EOICHOVLILGSLGQ-TVGAAPSPSELPGTGTAASSTK 234
Db 1539 -----ELQ--QOBQBPQCKTNTPOQKLPQLVSLPQLPPQAPPPQCPPLP-----QSSPS 1586
```



```

OY 235 PLPLPSPKPVOTSKTLASSSSSSSSSGAETPKO-----AFF 273
    ||: ||: ||
Db 1587 PSQLSHLPLKPLHTS-----TPQOLANLPOLIPYQCDCKCLAFPSFE 1629
OY 274 H-----LYHPLGSOHPFSAGVGRSHKPTPAPSPALPGSTDOLIASPHLAFPTTGLTAA 328
    ||: ||: ||
Db 1630 HMQEHQQLHFLSAQNOF-----IHPQLDRSLDMPEMLDPSNPLIAS 1672
OY 329 QCLGAA-----RGLE-ATASGLLKPKKSG-----353
    ||: ||: ||
Db 1673 QLLSGAIPQIPASSATSPSTPTMTLTKRLEKASASPG-----ENDSGTGGEPPORDK 1728
OY 354 -----ELSYGEVM---GPLEK-----366
    ||: ||: ||
Db 1729 RLKRTTPEQLEILYQKTLIDSNPTKMLDHTAHENYGLKRRVQVWFONTARERKKQFR 1788
OY 367 ---PGGRH-KCRFCAKVSGDSALQIHLRS---HTGERP-YKCNV-----CGNRTTGRG 412
    ||: ||: ||
Db 1789 AVGPAQAHRRCPFCALPKAKTALFAHIRSRHWEAKRAGYNLTLFSLMLDCCDGLQMKG 1848
OY 413 NLKVFHRRHREKYPVQNPVHPREHLDYVITSSGLPYGMSVPEPEKAEETATPGGYER 472
    ||: ||: ||
Db 1849 DI-----FDGTSFS-HLPPSSSD-----GQGVPL 1871
OY 473 KPLVASTTALSTESLTLLSTSA-----GTATAPGLPAFN-KFVLMA-----VEPKKA 521
    ||: ||: ||
Db 1872 SP-VSKTWELSPR---TLSPSSIKVEGIEDFESPSMVNLFNDQTKLDNDCCSVNTA 1927
OY 522 DENTPSEGSASISGVAASSTATLMQLSKLMTSLPSWALLTNHFKSTGSPFLPLCARALG 581
    ||: ||: ||
Db 1928 ITDPTTGDEGN-----ADNDSATGATEETKSSSAPNEGTLKKAAMAMEYEDRLSSGLVS 1982
OY 582 ASPESTSKLQOLVEKIDRQGAVAVTSASGAPTTSAAPSSSASGPN-----629
    ||: ||: ||
Db 1983 PAPSFYSK-----EYPNCGTVDYSETSSILADPCS-PSFGASGSAGSGSDGRPGQKRF 2035
OY 630 -----OCVICLRVLSGPR-----ALR 645
    ||: ||: ||
Db 2036 RTQMTNLOLKLKSCFNDRYRTPLMECEVGLGNDIGLPRKYVQVWFQONARAEEKSKLSMA 2095
OY 646 LHYG--OHGGERP-FKCYVCGRAFSTRGNLRAH-FVGHKASPAARAONSCTPCKKFTNA 701
    ||: ||: ||
Db 2096 KHFGINGTSYEGRPTECTGLIKYSARLSVRDHFSSQOHSKVDITQISQLDKEKEYFDP 2155
OY 702 VTLOOHVAMHLGGQIPNGSTALPEGGAOENGSEOSTVSA-----GSFPOQSOQSPSE 757
    ||: ||: ||
Db 2156 ATVQOLMAQOELDIRIKANEVL---GLAAOQOQMFNDPTPLQALMLPTAYPALQGIIPVLL 2212
OY 758 EELSEEEDEDEEEDVD-EDSLAGRSGESGGEKAISVGDSEASGAEEVEGVAA 816
    ||: ||: ||
Db 2213 PGLNSPLPGFPSTNALTSPKPNLMGLPSTTVPSPLPTISGLPNKPSASLSSTPPAQA 2272
OY 817 ATA-GKENDSNEKTQOSSLPPRPDSDLOPQPMEO-----852
    ||: ||: ||
Db 2273 TWAMPQRPQOQOQOQOYOVOQPPPPAPRAPPTPOLPQOQOQKDXSEKVEKEKA 2332
OY 853 ----GSSEVLGKEEGEGRKPERSSPASALTPEGGATSVTLVEELSLQAMRKEP 902
    ||: ||: ||
Db 2333 HKGKEPPLPVPRKEGGEAPTATATISAPLPTME-YAVDPQOLALQALATSDP 2385

RESULT 24
US-09-764-864-954
; Sequence 954, Application US/09764864
; Patent No. US2002012753A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT223
; CURRENT APPLICATION NUMBER: US/09/764,864
; CURRENT FILING DATE: 2001-01-17
; Prior Application data removed - consult PAM or file wrapper
```

```

; NUMBER OF SEQ ID NOS: 1792
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 954
; LENGTH: 708
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (328)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-864-954

Query Match          4.3%; Score 228; DB 10; Length 708;
Best Local Similarity 20.0%; Pred. No. 0.00019;
Matches 183; Conservative 86; Mismatches 326; Indels 320; Gaps 40;

OY 108 PTWGPGRGEESGHFLVATGTAAGGGGLIASPKLGATPLPESTPAPPPPPPPPP 167
    ||: ||: ||
Db 13 PLMHGPRGTSQSAAMNASEGESFAGS---VOIP---GGTVVLVLP-----55
OY 168 PGVSGHILNPLLELRLVLOQHQHQMOMTEQICRQVLLGSGQTVGAPASPELPCT 227
    ||: ||: ||
Db 56 -----DIHIC-----GICKQ--QPNNDATFAHKOSGCOLTGT 86
OY 228 GTASSTKPLPLPSPKPVQTSKTLASSSSSSSSSGAETPKQAFH---LYHPLGSOHP 284
    ||: ||: ||
Db 87 SAAPSTVQVSEETVPATQOTTTRITSETQIT-VSAPDEFEHGYOTVLPRESNEN 145
OY 285 FSAGV-----GRSHKPTPASP---ALPGSTDOLIASPHLAFPTTGLLAOCLGAA 334
    ||: ||: ||
Db 146 QTAIVLSLPAKSRTKKPTTPPAOKRLNCYCPG-----COFKTYGM-----186
OY 335 RGLEATASPGLLKPKNGSGLSYGEVWGPLEKPGRHKRCACAKVSGDSALQIHLRSHT 394
    ||: ||: ||
Db 187 KOMERH-----LKIHTG-----DKP---HCCEVCGKCFSRKDKIKTKHRCHT 225
OY 395 GERPYKCNVCGNRFETRGMLKVFHHRREKYP-HVQMPNHPVREHLDYVITSSGLPYGMS 453
    ||: ||: ||
Db 226 GVKPYKCKCTDYAAADSSSLMKHLRIHSDERPFQCOICPY-----265
OY 454 VPPEKAEETATPGGVERKPLVASTTALSTESLTL-LSTSAGTATAPGLPAFNKFVLM 512
    ||: ||: ||
Db 266 -----ASRNSOLIVLHRSHTGAPQCMWLSAKFKTS 298
OY 513 KAYERPNKADENTPSEGSASISGVAASSTATLMQLSKLMTSLPSWALLTNHFK---STG 569
    ||: ||: ||
Db 299 SDLKRRMRVHSGEKPRK-----CEFCNVKCTMK--GXIKSHIRIHNQSN 340
OY 570 SFPLPLCARALGASPESTSKLQOLVEKIDRQGAVAVTSASGAPTTSAAPSSSASGPN 629
    ||: ||: ||
Db 341 NFKCPHC-DFLGDSKATILKRHSRVHQ-----SEHPE 370
OY 630 OCVICLRVLSGPRALRLHYGOHGERPFCVKVCGRASTR--GNLRAPH-----676
    ||: ||: ||
Db 371 KCSECSYSCSKALRIHERIHCTDRPFCKNYC--SPDTKOPSNLSKHKHKGDMVKTE 428
OY 677 -----VGHKASPA-----ARAQNSCPICOKKFTNAVTLQOVRMH-----L 712
    ||: ||: ||
Db 429 ALERKDTGRQSSQVAKLDKKSFFHCIDCASFHREDSLRSHKROHSEYSEKNSDVTYL 488
OY 713 GGOI-PNGGTALPEGGAOENGSEOSTVSGAGSFPOQOQOQSPDEELSEEEDEDEE 771
    ||: ||: ||
Db 489 QFOIDPSKQPATPLVYGHLD-----VPLQPSQVP---QFSE-----521
OY 772 EEDVDDEDLAGRSGESGGEKAISVRGDSSEASGAEEVEGYTAAATAGKENDSNEK--T 829
    ||: ||: ||
Db 522 -----GRVKIIVGHQ-VEQANTIYQAAAANAIVPALVAQNPDELPGNSRLQI 569
OY 830 QOQSSLPPLPPPD-----SLDOPPM-----EOGSSGVL-----GKEEGEKPR 869
    ||: ||: ||
Db 570 LKQVSLIAPQPSRCSEEAAMTQPAVLLTTHBQDQATLHQTLLPTASGGQPEGSGNGT 629
OY 870 SSSPA-----SALTPEGGATSVTLVEE--LSIQEAMRKEGESSRRACEVCGQA 917
```



```
DB 630 FITSSGTCITDCEGLNALIOGCTA-EVTIVSDGGQNIUAVTATPAVPSSSSQ-----QE 682
QY 918 FPGQALIEHQKTHP 932
DB 683 LPKQTVYSIIQGAHP 697

RESULT 25
US-09-764-864-947
; Sequence 947, Application US/09764864
; Patent No. US20020132753A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT223
; CURRENT APPLICATION NUMBER: US/09/764,864
; PRIOR APPLICATION DATA REMOVED - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1792
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 947
; LENGTH: 323
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-764-864-947

Query Match
Best Local Similarity 24.4%; Score 227.5; DB 10; Length 323;
Matches 90; Conservative 45; Mismatches 133; Indels 101; Gaps 16;

QY 365 EKPGRHRCRCFCARVFGSDSALOHLRSHTGERPYKCNYCGNRFRTTSGNLKVFHRRREK 424
DB 11 EKP---HRCILCPFASAYERHLEAHMRSHTGKPYKCELCFSRDSRNLNLS---HHRRRK 64
QY 425 YPHVQNM-----RPHVPEHLDYVIT-SSGLP 449
DB 65 HKMVPKIGTRSSLSKKMWGLQKTSNLGYSRRALINLSPPMVVOKPPYLDNFTHEIP 124
QY 450 YGMSVPEKAESEA-ATPGGVERKPLVASTALSTESITLSTSGTATAPGLPAFNK 508
DB 125 ---NIQTDSTESMAKTTPTGGLPRD-----OELMVDNPLNQLSTLAGQLSS--LPP--- 171
QY 509 FYLMKAVEPKNKADENTPPGSEGSALISGVAESSTATLMQLSKLTSLPFWALLTNHFKST 568
DB 172 -----ENQNPASPDVYPCDEKPFM-IQPSQTQAV--VSAVASASIPQ-----SSS 213
QY 569 GSFPPLPCARLALGASPSSETSKLOQLVEKIDROGAVAVTSAASGAPTTSAAPSSSASGCP 628
DB 214 PTPSEP-----RPSHSQRNYSFV-----AGPSEPSAHTSTPSIGNSQPTPAPALP 260
QY 629 NQ-----CVICLRVLSCPRALRLHYGONGGERPFCKVCGRAFSTRGNLRAHFVGHKA 681
DB 261 VQDPQLLHHQCHDCMYFADNLIYTIHMGCHGYENFPQCNICGGCKKKNKYFACHF----- 315
QY 682 SPARAQNS 690
DB 316 ---ARGQHN 321

RESULT 26
US-09-764-864-1095
; Sequence 1095, Application US/09764864
; Patent No. US20020132753A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT223
; CURRENT APPLICATION NUMBER: US/09/764,864
; PRIOR APPLICATION DATA REMOVED - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1792
; SOFTWARE: PatentIn Ver. 2.0
```

```
; SEQ ID NO 1095
; LENGTH: 323
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-764-864-1095

Query Match
Best Local Similarity 24.4%; Score 227.5; DB 10; Length 323;
Matches 90; Conservative 45; Mismatches 133; Indels 101; Gaps 16;

QY 365 EKPGRHRCRCFCARVFGSDSALOHLRSHTGERPYKCNYCGNRFRTTSGNLKVFHRRREK 424
DB 11 EKP---HRCILCPFASAYERHLEAHMRSHTGKPYKCELCFSRDSRNLNLS---HHRRRK 64
QY 425 YPHVQNM-----RPHVPEHLDYVIT-SSGLP 449
DB 65 HKMVPKIGTRSSLSKKMWGLQKTSNLGYSRRALINLSPPMVVOKPPYLDNFTHEIP 124
QY 450 YGMSVPEKAESEA-ATPGGVERKPLVASTALSTESITLSTSGTATAPGLPAFNK 508
DB 125 ---NIQTDSTESMAKTTPTGGLPRD-----OELMVDNPLNQLSTLAGQLSS--LPP--- 171
QY 509 FYLMKAVEPKNKADENTPPGSEGSALISGVAESSTATLMQLSKLTSLPFWALLTNHFKST 568
DB 172 -----ENQNPASPDVYPCDEKPFM-IQPSQTQAV--VSAVASASIPQ-----SSS 213
QY 569 GSFPPLPCARLALGASPSSETSKLOQLVEKIDROGAVAVTSAASGAPTTSAAPSSSASGCP 628
DB 214 PTPSEP-----RPSHSQRNYSFV-----AGPSEPSAHTSTPSIGNSQPTPAPALP 260
QY 629 NQ-----CVICLRVLSCPRALRLHYGONGGERPFCKVCGRAFSTRGNLRAHFVGHKA 681
DB 261 VQDPQLLHHQCHDCMYFADNLIYTIHMGCHGYENFPQCNICGGCKKKNKYFACHF----- 315
QY 682 SPARAQNS 690
DB 316 ---ARGQHN 321

RESULT 27
US-09-764-864-893
; Sequence 893, Application US/09764864
; Patent No. US20020132753A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT223
; CURRENT APPLICATION NUMBER: US/09/764,864
; PRIOR APPLICATION DATA REMOVED - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1792
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 893
; LENGTH: 361
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-764-864-893

Query Match
Best Local Similarity 17.9%; Score 226.5; DB 10; Length 361;
Matches 108; Conservative 43; Mismatches 161; Indels 291; Gaps 15;

QY 365 EKPGRHRCRCFCARVFGSDSALOHLRSHTGERPYKCNYCGNRFRTTSGNLKVFHRRREK 424
DB 9 ERP---PECEBCGAKFRKSTVIYIHRTHHTGKPYECNCGKAFYQKSNLIYHQKTH--- 62
QY 425 YPHVQNMHPVPEHLDYVITSSGLPYGMSVPEKAESEAATPGGVERKPLVASTALSA 484
DB 63 ----- 62
QY 485 TESTLITLSTAGTATAPGLPAFNKFVLMKAVEPKNKADENTPPGSEGSALISGVAESSTAT 544
DB 63 TGKTYECTKCG-----ESFI-----OKLD----- 82
```



```

QY 545 LMOGLKMTSLPSWALLTNHFKSTGSEPLPLCLARALGASPSETSKLOOLVEKIDROGAVA 604
Db 83 -----LIHSTHTG----- 92
QY 605 VTSAAAGAPTTAPAPSSSSASGSPNOVCILRYLSCPRALRLHYGOGHERPFCVKGR 664
Db 93 -----KRPHECNECKKTTSDKSTLLIHORTHTGEEKHNKCTECK 131
QY 665 AFSTGNLRANHFVGHKASPARAONSCPICOKKFTTNAVTLQOHVBNHLGGQIPNGTALP 724
Db 132 SENEKSTL-----IVHORTHTGEEKPYEEDVCGKFTQKSNLTVHORTHSG----- 176
QY 725 EGGGAQNGESBQSTVSGAGSFQQQSQSQSPREBELSEEEDEEDEDYDDEDSLGR 784
Db 177 -----EKPEECNECEKAFQSKYLMH 198
QY 785 GSESGGEKAIIVRGDSEASGAEVEGVTAAMATAGKEMDENKTKTOOSLPPPPDSL 844
Db 199 GRGHGGEKPY-----ECNCEKAFQSKSYL-----II 225
QY 845 DQPPHREGSSSVLGKGEKGEKERRSSPASALPRREGATSVTLVEELSLQAMKREPE 904
Db 226 HORTHTEE-----KPYKCECGKAFR-----EKSKLLIHORTHTGE 261
QY 905 SSSRACAEVCGQAFPSQAALEEHOKTKPRKEGPLFTVCVRQOGLERATLKHMLLANHOV 964
Db 262 KPYE--CPYCWMAFQSKQSLLIHORTHTGEPKPYACTSCGAFKREKSTPTTVNO-RHTHGE 317
QY 965 QPF 967
Db 318 KPY 320

```



```

      9  SCDSSLEKEFLG--APVGSVSTPNSSQSSPSRSLANSIYEMVSDSESRRLG----- 61
      y 286 SAGVGRSHKPTPAPAPALPGSTIDOLIASPHLAPPSTTGLAAQCLGARGLEATATSPGL 345
      b 62 -----PBERLLDKDDSVYEDSLSEP-----LGVC----- 86
      y 346 LKPKNGSELGYGEVWGPPEKPGGRHKRCRCAKYVGSDSALQIHLRSHTERPYKCNVCG 405
      b 87 -----DGSGEPHSP--GGIRLPNGKLKCDYCGMVCIGPNVLMVHKRSHTERPFCNNOGG 140
      y 406 NRFTGRNLKVFPHRREKYP-----HYQMPHPVPEHLDYVTTSS--GLPY----- 450
      b 141 ASFTQKGNLRLRIKLHSGEKPFCPCPCNYACRRDALTLGHLRTHSVSSPTVGKPYKCNVC 200
      y 451 GMSVPEKAEFE-----AATPGCGVEKRPVASTATLSTESTLL 491
      b 201 GRSTYQOQSTLEHKKECHNYLQSLSTDAQALTGQPGDEIRLDEWPPDSMLHSTERPTTI 260
      y 492 STSAGTATAPGLPAFNKEVLMKAV-----EPKNKADENTPPGSEGS 532
      b 261 DRLANSLTKRKRSSTPQKPYGEKQMRFSLDLPYDVNAGSYEKDVELVAHGLPEGFGS 320
      y 533 -AISG-----VAE-----SSTATLMQLSKMTSLPSSWALLTNHFKSTGSP 572
      b 321 LAFTGTEHLRLPLRPPTNCISLTPVISSVYTOMQIPSRLELP-----SREAGEGP 373
      y 573 L-----PLCARAL-----GASPSETSKLOQLEK--IDROGAVA----- 604
      b 374 EDLGDGGLULYARAGSLTDPGASPSNGCQDSTDESNEHEDIGGVLSLPQGPPTPTI 433
      y 605 VTSASGAPTTISAPAPSSS---ASSGPN-----QCVCILRVLSCPRAL 644
      b 434 VVGRHSPYAKEDPKPQEGELRLCTPGSPKEVLRYVSGEGEPAKFKCHC--RIILDHVM 492
      y 645 -RLHYGAGGEPEFKCKYCGRAFSTRGNLRAHFV--GHKASPARAONSCPT 693
      b 493 FTIMHGCHGRFPRFECNICGYSDRYEFSFHYRGHEKV-----GSCRI 537

RESULT 33
IS-09-083-290-2
Sequence 2, Application US/09083290A
Patent No. US20020023277A1
GENERAL INFORMATION:
APPLICANT: Stunlmann, Heidi
APPLICANT: Xiong, Jing-Wei
APPLICANT: Taudman, Mark B.
TITLE OF INVENTION: VASCULAR ENDOTHELIAL ZINC FINGER 1 GENE
TITLE OF INVENTION: AND PROTEIN AND USES THEREOF
FILE REFERENCE: 31200
CURRENT APPLICATION NUMBER: US/09/083,290A
CURRENT FILING DATE: 1998-05-22
NUMBER OF SEQ ID NOS: 13
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 2
LENGTH: 506
TYPE: PRT
ORGANISM: Mus musculus
IS-09-083-290-2

Query Match      4.2%; Score 222.5; DB 10; Length 506;
Best local similarity 17.5%; Pred. No. 0.00025;
Matches 133; Conservative 72; Mismatches 231; Indels 325; Gaps 22;

      y 191 QIHOMQTEQICROYLLIGSLGOTVGAPASPSSELPGTGTASTTKPLPLPFSPIKPVQTSK 250
      b 11 QAHSHAHQQAQDAQNSLRLPLSSAVERP-----DQKPLRLPIPTQKQAPE 57
      y 251 TLASSSSSSSSSGAFTPKQAF-----FHLVHPLGSHPPSAGVGAGSHKPTP 298
      b 58 TLADAIQIKK-----EKPKTSFVCTYCSKAFRDSYHL-----RRHOSCHNGIKILVSRRAK 107
```

```

      y 239 APSPALPGSTIDOLIASPHLAPPSTTGLAAQCLGARGLEATATSPGLKPKNGSELSTYG 358
      b 108 TPPTTVP-----LISTIADSSRTSLVST--IAGILSTVITSSSG-----TNPSSASAT 155
      y 359 EWMGP--LEKPG--GRHKCRFCAKYVGSDSALQIHLRSHTERPYKCNVCGNRPTTNG 412
      b 156 AMPVPQSVKRPSPKPVKNHACEMCGKAFRDYVHLNRHKLSSHDEKPFECPIQNRFRKRD 215
      y 413 NLKVFHRRREKYPHYQMPHPVPEHLDYVITSSGLPYGMSVPPKAEDEATPGGGER 472
      b 216 RMTYHRSHE-----GGITK 230
      y 473 KPLVASTALSTESLTLSTAGTATAPGLPAFNKEVLMKRAVEKKNKADENTPPGSEGS 532
      b 231 ----- 230
      y 533 AISGVAESSTATLMQLSKMTSLPSSWALLTNHFKSTGSPFLPLCARALGASPSETSKLOQ 592
      b 231 ----- 230
      y 593 LVEKIDROGAVAVTSAASGAPTTAPAPSSSASGPNOCVICIRVLSGPRALRLHYGQ--H 651
      b 231 -----PYTCSYCGKGFSPRDHLSCYKHYH 255
      y 652 GGERPFCKYCGRAFSTRGNLRAHFVGHKASPARAONSCPTCOCKFTNAVTLQOHVRMH 711
      b 256 STERPEFKCQCTAAFAFKDLRTHMVHRE-----GKVSCHNIC--GKLISAAYITSHLKT 308
      y 712 LGGQIPNGTALPEGGAQENSGEOSTVGAGSFPOQSOQSPSEEEELSEEEDEDEE 771
      b 309 GOSQISINMTC-----KQISIKTCMSEETSNQKQOQOQ-----QOQOQOQOQ 349
      y 772 EEDVVD-----EDSLAGRGSESGEKAISYRG-----DSEERAGAE 807
      b 350 QQHTVSWPQKQYETLRLWEAVAKRKKEANLQOTSTAATTPVTLTTPFNITSSVSSGTM 409
      y 808 EEVGVYAAATAGKENDSNEKTTQOSSLPPPPP--PDSLQPOPMEDGSSGVLDGKEBG 864
      b 410 SNPVTVAAMSRKSPVNVSSAVNTSPLAMTSPLTLTTPVNLPPVPT----- 456
      y 865 GKPERSSPASALTPPEGATSVTLVEELSLOEAMRKPEGS 905
      b 457 -APVNIARHVTTITSPNLTPTMTLAPLNT--AMR--PVES 492

RESULT 34
US-09-816-669A-12
Sequence 12, Application US/09816669A
Patent No. US20020137019A1
GENERAL INFORMATION:
APPLICANT: GARABEDIAN, Michael
APPLICANT: YANEJA, Samir
APPLICANT: HITTELMAN, Adam
APPLICANT: MARKUS, Steven
TITLE OF INVENTION: METHOD FOR SCREENING TRANSCRIPTIONAL COREGULATORY PROTEINS OF
TITLE OF INVENTION: TRANSCRIPTION FACTORS, AND ANDROGEN RECEPTOR TRANSCRIPTIONAL
TITLE OF INVENTION: PROTEINS AS TARGETS FOR ANDROGEN RECEPTOR-DEPENDENT DISEASES
FILE REFERENCE: GARABEDIAN-1.1A
CURRENT APPLICATION NUMBER: US/09/816,669A
CURRENT FILING DATE: 2001-03-26
PRIOR APPLICATION NUMBER: 60/225,618
PRIOR FILING DATE: 2000-08-15
PRIOR APPLICATION NUMBER: 60/191,768
PRIOR FILING DATE: 2000-03-24
NUMBER OF SEQ ID NOS: 20
SOFTWARE: PatentIn version 3.1
SEQ ID NO 12
LENGTH: 252
TYPE: PRT
ORGANISM: Human
US-09-816-669A-12

Query Match      4.2%; Score 221; DB 10; Length 252;
```


SEQ ID NO 957
LENGTH: 600
TYPE: PRT
ORGANISM: Homo sapiens
US-09-764-864-957

Query Match 4.1%; Score 218; DB 10; Length 600;
Best Local Similarity 18.8%; Pred. No. 0.0005;
Matches 129; Conservative 65; Mismatches 218; Indels 274; Gaps 22;

```
371 HRCRCQAKVFGSALQIHR-SHTGE--RPYKVCNCRNFTTGNLKVHFRHREKY-- 425
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 66 YNCTACNMDFRKPADQVHKSHLGNPAKAKHCLFCGETSETVEVLQCHITHSKRYNC 125
    :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 426 -----PHYOMNPHVPPEHLDY-VITSSGLPYGMSVPEKAEEDAATPGGVERKPLV 476
    :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 126 KFCSKAFHAILLEKHLREKHCVDATENGCTANG--VPP----- 163
    :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 477 ASTTALSATSESLTLLSAGTATAPGLPAFNKFLVMKAVEPKN--KADENTPPGSE---G 531
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 164 -----MATKKAEPADLQGLMLKNPEADNHSHEASEDDVDASEPMYG 203
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 532 SAISGVAESSTATLMQSLKMTSLPSMALTNHFKSTGSPFLPLCARALGASPESTSKLQ 591
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 204 CDICGAYTME-----VLQNH-----RLRDHNIRRGEDDGSR 236
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 592 QLEKIDROGAVAVTSAASGAPTTAPAPSSASSGPNOCVYICLRVLSCPRALRLHYGOH 651
    :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 237 KKAEEF-----KGSHKCNVCSRTFFSENGLREHLQTH 268
    :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 652 -GGERFFKVCGRAPSTRGNLRAHFVGHKAS-----PA 684
    :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 269 RGPAAHYMCPICGERFPSLLTTEHKYTHSKSLDGTGCRICKMPLOSEEFTIEHCQMPD 328
    :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 685 ABAQNS-----CPICQKKTFTNAVTLQOHVRLHGGQIPNGSTALPREGG-----AA 730
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 329 LR--NLUTGRVCVCMQTVISTLELKINGTFHMOKLAGSSAASSPNOGQLKLYKALCL 386
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 731 QENGSEOSTV-----SGAGSPQOOSQOP-----SPEEE 759
    :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 387 KEFRSKQDLVLDVNGLPYGLCAGCMARSANGOVGLAPPEPADRPAGLRCPESKVFE 446
    :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 760 LSPDEEEDDEEEDVDYDEDSLAGKSGSGEKALISVRGDSSEASGAEEVGTVAATA 819
    :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 447 SAEDLESHMVDHHDLPETS----- 467
    :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 820 GKEMDSNEKTTQOSSLPPPPPSLDPQPMEOGSSGLGKEEGKPERSSSPASALTP 879
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 468 -----GPRKGTQTSPPV-----RKTTYQCIKCQMTF 493
    :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 880 EGEATSVTLVEELSLQEMARKPPGESSSSRKACEVCGQAFPSQALAEHQTHPKE--GPL 937
    :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 494 ENEREIOIHVANHMIIEEGINHE-----CKLCNQMEDSPAKLLCHLIEHSEGMGT 544
    :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 938 FTCVFCRQGFLEATLTKKMLLAHHQ 963
    :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 545 FKCPVCTVEVOANKLOOHIFAVHGO 570
    :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
```

RESULT 40
US-09-974-298-118
Sequence 118. Application US/09974298
Patent No. US20020156263A1
GENERAL INFORMATION:
APPLICANT: Chen, Huel-Mei
TITLE OF INVENTION: GENES EXPRESSED IN BREAST CANCER
FILE REFERENCE: PA-0037 P
CURRENT APPLICATION NUMBER: US/09/974,298
CURRENT FILING DATE: 2001-10-04
PRIOR APPLICATION NUMBER: 60/238,331
PRIOR FILING DATE: 2000-05-10
NUMBER OF SEQ ID NOS: 194
SOFTWARE: PERL Program

SEQ ID NO 118
LENGTH: 516
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc.feature
OTHER INFORMATION: Incyte ID No. US20020156263A1 5665139CD1
US-09-974-298-118

Query Match 4.1%; Score 217.5; DB 9; Length 516;
Best Local Similarity 24.5%; Pred. No. 0.00045;
Matches 86; Conservative 29; Mismatches 135; Indels 101; Gaps 9;

```
QY 365 EKRCGRHKRCFKAKVFGSDSALQIHLRSHTGERPYKVCNCRNFTTGNLKVHFRHREK 424
    ||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 258 EKP--YRCNVCGKVFPHHISHLAQHRIHTGEKPYKNCCKGKVFSHKSSLVNHRIITGE 314
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 425 YPHVOMNPHVPPEHLDYVITSSGLPYGMSVPEKAEEDAATPGGVERKPLVASTTALSA 484
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 315 KPYKNCCKGKVFSHKSSLVNHRIHTGEK--PYKNE---CGKVFSRNSYLA----- 361
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 485 TESTLTLSTAGTATAPGLPAFNKFLVMKAVEPKKADENTPPGSGSASISGVAESSTAT 544
    :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 362 -----QHLIIHAGEKPYKCDKCAFSQNS----- 386
    :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 545 LMQLSKMTSLPSMALTNHFKSTGSPFLPLCARALGASPESTSKLQOLEKIDROGAVA 604
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 387 -----HLVQHHRIHTGEKPY-----KDECGKV- 409
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 605 VTSAAAGAPTTAPAPSSASSGPN--OCVYICLRVLSCPRALRLHYGOHGERPPKCKVC 662
    :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 410 -----FSQNSYLAYHWRIRHTGEKAYKCNCEGKVFGLNSLHNRKIHTGEKPEKCMKC 462
    :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 663 GRAFSTGNLRAHFVGHKASPARAQNCSPIQCKKTFTNAVTLQOHVHMHG 713
    :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 463 GKAFMSRSSLTNNHAIH---TGERHFKCNCEGKLFPRDINSYLVRHQRFHAG 509
    :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
```

Search completed: January 13, 2003, 15:23:20
Job time : 27.0149 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 13, 2003, 15:17:45 ; Search time 22.5336 Seconds
(without alignments)
4287.595 Million cell updates/sec

Title: US-09-988-117-1
Perfect score: 5277
Sequence: 1 MAHESRSRLGVPAGEPAE.....PSITSTGLSPFRKDDPTIP 1005

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database :
1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1291	24.5	1350	2	T30341 zinc finger protei
2	1290.5	24.5	1323	2	spalt protein - mo
3	1145	21.7	1061	2	Xsal-3 protein - A
4	777.5	14.7	1355	2	spalt protein - fr
5	745.5	14.1	1402	2	finger protein - f
6	727	13.8	1263	2	spalt-related prot
7	391.5	7.4	744	2	hypothetical prote
8	346.5	6.6	2282	2	DNA-binding protei
9	330.5	6.3	727	2	transcription regu
10	322	6.1	1615	2	ras-responsive ele
11	316.5	6.0	2232	2	hypothetical prote
12	313	5.9	1920	2	gene hindsiht pro
13	308	5.8	1891	2	hypothetical prote
14	307	5.8	1173	2	prockr2 - chicken
15	301.5	5.7	651	2	finger protein 2,
16	296.5	5.6	1350	2	finger protein - A
17	296	5.6	644	2	hypothetical prote
18	293.5	5.6	744	2	finger protein bow
19	293.5	5.6	1191	2	zinc finger protei
20	292.5	5.5	982	2	hunchback-related
21	291.5	5.5	1184	2	atrophin-1 - human
22	286	5.4	1184	2	atrophin-1 - human
23	281.5	5.3	624	2	DNA-binding protei
24	280	5.3	1060	2	homeotic protein z
25	279.5	5.3	1173	2	C-terminal domain-
26	277.5	5.3	485	2	finger protein MZF
27	277.5	5.3	594	2	ayliphorin gene-sp
28	277	5.2	654	2	finger protein ZNF
29	277	5.2	671	2	krueppel-type zinc

30	276	5.2	1042	2	A31591
31	275	5.2	803	2	S26823
32	275	5.2	2715	2	T13049
33	271.5	5.1	581	2	A19073
34	265	5.0	856	2	A35503
35	264.5	5.0	543	2	E88280
36	264.5	5.0	615	2	S06546
37	264.5	5.0	1051	2	A60191
38	264	5.0	428	2	finger protein (cl
39	264	5.0	595	2	UC7779
40	263.5	5.0	789	2	A39564
41	263	5.0	604	2	S05447
42	263	5.0	2688	2	I49477
43	261	4.9	553	2	S22954
44	260	4.9	595	2	G02075
45	260	4.9	1042	2	S41705
46	259	4.9	701	2	T14757
47	258.5	4.9	2578	2	A56922
48	258	4.9	710	2	I48668
49	257.5	4.9	636	2	I48689
50	257.5	4.9	675	2	S51037
51	256.5	4.9	2529	2	A56923
52	256	4.9	488	2	S47072
53	255	4.8	686	2	A34612
54	254	4.8	439	2	S06556
55	254	4.8	591	2	S65088
56	253.5	4.8	1154	2	A56242
57	253.5	4.8	2717	2	A34203
58	253	4.8	1117	2	UC4934
59	252	4.8	536	2	S06548
60	252	4.8	589	2	I38598
61	250.5	4.7	469	2	I38600
62	250	4.7	509	2	JH0501
63	248.5	4.7	496	2	T08674
64	248	4.7	582	2	S08686
65	247	4.7	908	2	T16057
66	246.5	4.7	727	2	G01792
67	246	4.7	794	2	S59069
68	245	4.6	2500	1	WMHUE2
69	244.5	4.6	3942	2	T42730
70	243	4.6	546	2	I4636
71	242.5	4.6	548	2	S42077
72	242	4.6	542	2	A54661
73	242	4.6	693	2	I37570
74	241	4.6	728	2	A54603
75	240	4.5	1186	2	T33754
76	239.5	4.5	625	2	D87793
77	239	4.5	411	2	S10245
78	238.5	4.5	540	2	B57785
79	238.5	4.5	1706	2	I84499
80	238	4.5	494	2	A42170
81	237.5	4.5	1320	2	JC5630
82	236	4.5	337	2	S06566
83	236	4.5	420	2	S06579
84	234.5	4.4	955	2	T00247
85	234.5	4.4	1561	2	T00248
86	234	4.4	614	2	JH0500
87	233.5	4.4	688	2	A56360
88	233.5	4.4	1124	2	JX0293
89	233	4.4	1585	2	T31611
90	232.5	4.4	477	2	A47236
91	232.5	4.4	580	2	A37107
92	232	4.4	594	2	T12488
93	232	4.4	1106	1	TVHUGL
94	231.5	4.4	497	2	JC5076
95	231.5	4.4	816	2	S05548
96	231	4.4	728	2	A48830
97	230.5	4.4	1114	2	I50222
98	230.5	4.4	3938	2	T42761
99	229.5	4.3	732	2	S47073
100	229.5	4.3	1043	2	A56037

transcription regu
zinc finger protei
eyelid - fruit fly
HSN motor neurons
B-lymphocyte-induc
finger protein (cl
oncogene Evi-1 - h
finger protein (cl
krueppel-associate
transcription repr
finger protein glia
alpha-A-crystallin
finger protein zip
transcription repr
Evi1 protein - hum
hypothetical prote
transcription fact
zinc finger protei
gene NK10 protein
zinc-finger protei
transcription fact
finger protein HZF
S47072
zinc finger protei
finger protein (cl
finger protein XFO
E-box-binding repr
DNA-binding protei
delta-crystallin/E
finger protein (cl
zinc finger protei
zinc finger protei
zinc finger protei
probable-finger pr
finger protein ZFP
hypothetical prote
transcription fact
Z13 protein - mous
HIV-EP2 enhancer-b
Bassoon protein -
DNA-binding protei
zinc finger protei
zinc finger protei
zinc finger protei
transcription fact
O/E-1-associated z
protein C27A12.2
finger protein, te
zinc finger protei
zinc finger protei
zinc finger protei
TCOF1 protein - mo
finger protein (cl
finger protein (cl
zinc finger protei
zinc finger protei
zinc finger protei
hypothetical prote
zinc-finger protei
spermatogenesis pr
hypothetical prote
transforming prote
myc-associated zin
gap protein hunchb
probable transcrip
deltaE1 - chicken
Bassoon protein -
finger protein HZF
DNA-binding protei

QY	216	GAPRSPBL-----	PGGTASTRPLRLPSP-----	IKPVOT	248
Db	213	GRLPKPSAAGNMNTSVOLGTLPHALLOLSAGPATASGSGTLPAEFDPOHLSOPASG			272
QY	249	SKTLASSSSSSSSGAEETPKQAFPHLYPLGSOHP-----	FSAGVQGRSHKP-----	TP	298
Db	273	TSTCCSTSAAPRDSGA-----	HPACSTGPAPGVAANAASSTGVNAYQPOAHSAP		321
QY	299	-----	APSPALPGSTDOGLASPHLAFPPSTGLLAOOLCGAARGLSATASFC		344
Db	322	PALGGRPLSSASNLPRNLPLOT-----	SSSSVIREPNPLVSTIA-----	TAMALDPLISA--	370
QY	345	LLKPRNDS--GELSYCEWVGPRLEKPEGGRKCRCAVFGSDSLQTLNBSHTGERPKCV			403
Db	371	LKMHKGRPNVSVFERKASADEPFKHKCRCAKFGSDSLQTLNBSHTGERPKCN			430
QY	404	CGNRPETGNLKVHNRHREKYPHYOMNRPVRENLYITSAGLRYGASVPEK-----			458
Db	431	CGNRSPTGNLKVHNRHREKEXPHYOMNRPVREYLIDKPCSGIYUGSLPREKCVTTM			490
QY	459	AEEBATP-----	GGVERKRLVASTALSTATESLTLSS-----	TSAGTAPGL--	503
Db	491	LDSKVLPTVPSTVGLDRLPTVPGTHNYTDSSTIPVSSRQPRSPASSECTSLSGLNN			550
QY	504	-----	PAFNKFLMKAVEPKKABENPRGSE--	GSAISGAESTATIMQ	547
Db	551	TESGITVBPRESQPLLGSRPLTKA--	EVSLPCTSTTGAPRVUGGVSGTLPSTAAITVD		609
QY	548	LSKILMTSLPSMALLTNHFKSTGSPFLPLCARALGASPSSETSKLOQJLEKIDROGAAVNTS			607
Db	610	SACSTLGSBGLPRAVDQFKA--	QFRPGGLDSM--	QTSKSLQOLVENTDKR-----	658
QY	608	AASGAPTTSAAPASSASSGPNOCVILRYLSCPRALRYHNGONGERPRKCVGSAFS			667
Db	659	-----	MTDPMQCVICHRYVLSGOSALKMHYRTHNTEBERPRKXICGRAFT		701
QY	668	TRGNLRANRVGHKASPAALAAQNSCPTCKKFTPAVTLQOONVHNLGGOTNGTALPREG			727
Db	702	TGNLKTHTGVHGRKPRLEKRVONHSCPTCKKFTPAVTLQOONVHNLGGOTNGTALPREG			759
QY	728	GAA-----	QENSSSEOST-----	VSGAGSFPQOOSQ	752
Db	760	QEAAMADLPFDEKNAETLSSFPDDIDENSMEDSELSKDTASDSKPLSLVSSGCP-----			814
QY	753	QSPPE-----	ELISEEEEBEDEEBEDVYDEDSLACRG--	SE	787
Db	815	DSPPSVISSIALALENQMKIDSVMNQOOLANKLSEVNGSGESGDRLSNDSSSAAGLESR			873
QY	788	SGGEKAISVRGDSSEASGADEEVGTVAAATA--	GKENDSNEKTTQOSSLRPPRPBDSLQ		846
Db	874	SAGSPRLSSSSSSQALSPHNSNGESRFRSPGLGHOEDPELTKTERLDSPP-----			926
QY	847	POPMEGSGSGVLGKEEGGKP--	BRSSSPASALPRREGATSVTLVEELSLQEDAMKEGES		905
Db	927	PGRGNGALDLTAG--	HGRPLIKEAPRSL-----	FLSRENGKC	965
QY	906	SSRKACEVCGAQRPSQALAEHQKTHPKEGRLTYCVFCROGFLERATLKHMILANHOVQ			965
Db	966	AS--TVGCVGCKPRFACKSALEIHYRSHTKE--	RREVYTCVCRGCGSTMGLKQILTL--	HKLK	1021
QY	966	PFAHPHQPUNIALSLVPGCSPTITSTGSLFPP	997		
Db	1022	EL-----	PSQVFDENFTLGPBHSHTPSLASSPAP	1049	

RESULT 3
JC7116
Xsal-3 protein - African clawed frog
C:Species: *Xenopus laevis* (African clawed frog)
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 11-May-2000
C:Accession: J07116
R:Onuma, Y.; Nishinakamura, R.; Takahashi, S.; Yokota, T.; Asashima, M.

Biochem. Biophys. Res. Commun. 264, 151-156, 1999
A:Title: Molecular cloning of a novel Xenopus spalt gene (Xsal-3).
A:Reference number: JCT116; MID:J99458636; PMID:10527856
A:Accession: JCT116
A:Molecule type: DNA
A:Residues: 1-1061 <OND>
A:Cross-references: DDBJ:AB030827
A:Experimental source: egg, neural tube
C:Genetics:
A:Gene: Xsal-3
C:Keywords: egg; zinc finger

Query Match	21.7%	Score 1145;	DB 2;	Length 1061;
Best Local Similarity	31.5%	Pred. No. 1.2e+47;		
Matches 324;	Conservative 140;	Mismatches 347;	Indels 218;	Gaps 36;

QY	18	PALEGGDAS-----BEDRPVCAKCCAOPTDPTPEFLAHONACSTDPRVWYJIGGOEJPN	71
Db	26	PSSVYGERVAKRCMEETH--ICEKCCAEFFELSDLEHKKSCKTPTVLLMNDG--EGAM	82
QY	72	NSSASSEPREPEGHNNQVMDTEHSNPPDSSGVPTDPTWOPERKGEDESSGHFLVAATGTA	131
Db	83	SHDGYTEESPEEG-----AAEESEVPAVDSYQPKRSSVPEKEEKYDSNKTVPQNSK	136
QY	132	ACGGGGLIASPKLGATPLPESTPAAPPPPPPPPPPPGGSHLN---IPILIELRVLQ	188
Db	137	TNGLG--YVPKTLSTNTNTLTQINSTKVAVNOHASDGVATSATNPNAIPILIQVLQ	194
QY	189	ONOHOMOMTEOICROVULLGSLGTVGAPAS--PSELPGTASTRKPLPLESPIKPV	246
Db	195	QOOLQOTLQTEOIRIOIAM-----APNSLHPS-----IAAIDPLKALGHLISQ	240
QY	247	QTSKTLASSSSSSSSSGAETPKQAEFHLYHPLGSOHPSAGVGRSHKPTPADSPALPG	306
Db	241	LSAAVALIGOKACTOGLSTLESKO-----SKRPHSNVAM-----PAGTVP--DAL	284
QY	307	SHDQLIASHPLAPRTTGLLAOCIGARGLENTAP--GLIKPKNGSELXYGEVWGPLE	365
Db	285	TTSCLKOEPLGLTNVGRFPNPALPHSPGIIIFQNTINALDPESK--KIKVAFPTYTTEA	343
QY	366	KKGG-----RHKKRFGCAKAVGSDSALOIHURSHTERPERYKCNVCNRPRTTGNLKVHFR	420
Db	344	KGNENQQLFRHKKCFQCGKVGXGNSALQIHLRSHGERPYKCNIGCNRPRTTGNLKVHFOR	403
QY	421	HREKTYHVQMNPRVPEHLDYVITSSGLPFGMSVPRPEKAEFEAATPGGVERKPLVAST	480
Db	404	HKKYTHIKMNPVPVEHLDVNPRTSSIPGMSVPLDESNLIADTKSG-----LT	453
QY	481	ALSATESLTLSTSAGATAPGLPAFKFVLMKAVEPKNADEMTPESEGAISG---V	537
Db	454	GLPSAINLSGLTESV-----LGAF-----PLN--MOSRNPSESEGSVGAHV	495
QY	538	AESSTATLMOLSKLMTSLPSMALLTNNHFKSTGSPPLPICARALGAP--SETSKIQOLVE	595
Db	496	QESGTD-----OSNESPVS-----GSSDQSEETTKLOOLE	527
QY	596	KIDROGAVNTSAASGAPTTSAAPASSASGCPNOCVTLCEVLSRPLRLRYHGOHGER	655
Db	528	NLDKNG-----SETNECLICHVLSPPSSILKMRYHTHTGR	563
QY	656	PERCKVCGRAEFTSGNLRHAFVGHKASPAABAONSCPICOKKFTNAVTLQOHVRYMLGSO	715
Db	564	PFCKICIGRAPFSTKNTKHGYHVRANPLKLOHSCPICOKKFTNAVTLQOHRIMMGSK	623
QY	716	IPNGGTLPEGGGAQOENGSRQIVSGA--GSFPOQ-----SOQSP	756
Db	624	IPN-----TPVSEESDIDISMDXKNGELNNSFTDENLQDIDMEDDELAENASGSKPFP	679
QY	757	EEELSEE-----EEDEDEEEDYTDDESL-----AGRSESQGEKALSVRGDSEBASA	806
Db	680	HSFTRAESPAMQFSTGGOQDKPVTPLPSALNLOQRQSVKSENGSLESDELTDNDSSVMDQ	739
QY	807	EEEVGYVAAATACKENDSNKTTQOSSLPPP-----PPDSLQOPQPMEDGSSG	856


```
Db 740 EYPRTGSPQTSEARTEFPTNSQSDSNASKSPSYNGLDLGLMLSKDEHSGNSLNPQDGG 799
OY 857 VL-----GG-----KEEGKRPERS-----SSPASALTPREGATSVTLVEELSLQEA 897
Db 800 ALDLINGGFARKIKKEEFGDLHONGERGLPNLVGAPRALI--KMEVSSDMSAGATQYLGP 857
OY 898 MKREPG-----ESSSRKACEVCGOAFPSQALAEHQKTHPKEGSPLETCVFCRQGL 948
Db 858 PNLSPGLNPLIVGQRRSAKQHCIMCGKNFSASALQIHERHTHGKRP-FACITICGNAFT 916
OY 949 ERATLTKKH 957
Db 917 TKGNIKVVH 925

RESULT 4
S40022
spalt protein - fruit fly (Drosophila melanogaster)
C/Species: Drosophila melanogaster
C/Date: 19-Jul-1996 #sequence_revision 26-Jul-1996 #text_change 21-Jul-2000
C/Accession: S40022
R:Kuehnlein, R.P.; Frommer, G.; Friedrich, M.; Gonzalez-Gaitan, M.; Weber, A.; Wagner-Ber-
EMBO J. 13, 168-179, 1994
A>Title: spalt encodes an evolutionarily conserved zinc finger protein of novel structure
A/Reference number: S40022; MUID:94139659; PMID:7905822
A/Accession: S40022
A:Molecule type: DNA
A:Residues: 1-1355 <KUP>
A:Cross-references: EMBL:X75541; NID:g414106; PIDN:CA53229.1; PID:g2598394
A>Note: The authors translated the codon GAC for residue 51 as Ala
C:Genetics:
A:Gene: sal; spalt
A:Cross-references: FlyBase:FBgn0004579
A:Introns: 51/1; 1329/3 1355/2
C:Keywords: zinc finger

Query Match 14.7%; Score 777.5; DB 2; Length 1355;
Best Local Similarity 23.6%; Pred. No. 6,6e-30;
Matches 308; Conservative 129; Mismatches 395; Indels 471; Gaps 45;

OY 68 ENPNSSASSERPREGHNNPDVMTHEHN--PPDSGSSVPTDPTGPRRGEESGHFL 124
Db 106 ENDIKSEAKSEIEPVEDNNRNVAMTKPSSEREBNASGSMSPV--AFASAE----- 157
OY 125 VAATGTAAGGGGGLIASPRLGATP-----LPPESTAPRPPPPPPPPVGV 171
Db 158 -AAT-----A-----EKTPKEKEKDVENVDMPEAPSSAVPSTEVTLPGAG 199
OY 172 SGHLNPLILEELR-----VLOQRQIH 193
Db 200 A-----PYTLEAIQNMQMAIAQFAAKTIANGSNGADNEAMKQALFLOQLFNLOQQLF 254
OY 194 OMQMEQICROVLL-----LGSLGQTVGAPASPS 222
Db 255 QIQLIQLOLSQALMQAKQEDTEEDADQEDQEDQEDTDTVEEERIALMELRQKAEARMA 314
OY 223 ELPGTASSTKPLPLFPIKPVQTSKTLASSSSSSSSSGAGPKQAFPHLYHPILGSO 282
Db 315 E-----AKARHILNAGVPLR-----BSSGSPAESLKRREHDH-----ES 350
OY 283 HPFSAGVGSRHKPPRA-----PSPALPGSTD--OLIASPH--LAFSTGLTAA 328
Db 351 QPNRTSLDNTHTKADTAQOALAKLKLEMEENTPLRPGSLASSITINHDDLPRENSIDLQK 410
OY 329 QCLGAARLEATASGLLKPKNSSGELSYGEVGR---LEKPGGRHKCFCAKAVGSDSA 385
Db 411 R---ADEVLSASQIL--ANSWADPFAFEKSGEGKGRNPFKKHRCRYCGKVGSDSA 465
OY 386 LQIHLSHTGEPRYKCNVCGNRTTRGNLKVHFHHRREKRYHGYMNPVRYEH----- 438
Db 466 LQIHRSHTGERPRKCNVCGSNFTTKGNLKVHFRQHAQKPRPHVPMNATPRIDEHMDKFNRP 525
```

```
OY 439 -LDVYITSSGLPYGMSVPRPEKAEAEAPPG--GVBR--KPLVASTTALSATESLTLLST 493
Db 526 LLDQMSPTDSSPNHSPAPRLPGSAPASFPAPRGLONLVRRPEHILKSLGAAPHQYEPQ 585
OY 494 SAGT---ATAPGLPAFNFLVLMKAVERPKKADENPRPGSESAIS-----GVASST 542
Db 586 ELPTDLAKRPSQDLDEDEQVANKERVEEKDQREHEDMAECSEPEEPRLPLEVRIKERY 645
OY 543 ATLMOISK-----LMTSLPSMALLT--NHRKSGSPF--LPLCARAL---GASPSFT 587
Db 646 EEQEQVKQEDHRIEPRRTSPSSSEHRSPHNHRSHMGVPPVYQIQPALMHQSSPSGQ 705
OY 588 SKIQQL-----VEKIDROGAAVYMSAAGAPTSAPAR--- 620
Db 706 SHLDLPLEPQOLPREDEFAERFPLNFTYAKMLSPENHSVRSFAGALRPGVPPPHH 765
OY 621 -----SSSASSGPNQ 630
Db 766 PHHMAISPFENPIKHEMAALLPRHSHNDSNMFNTEVSTCEIMKLKELMKKKKISDPNQ 825
OY 631 CVTCLRVLSCPRALRLHYGOHGERPEFKCKYCGAFSTRGNLRAHFVGHKASPAARAONS 690
Db 826 CVYCDRLVLSCKSALQMHYRTHTGERPEFKRCIGRAFETTKGNLKTMAVHKIRPRMRFHQ 885
OY 691 CRTCKRKTMAVTLQOHVBMILG----- 713
Db 886 CPVCHKKYSNALVLDQITRLTGTERTDLPQIOAAEIRDPSPSMRGMFMFPAAAFH 945
OY 714 -GGIPNGTALPREGGAAGAENGSDQTVSAGSFPOQSOQSPSEBELSEE-----EE 765
Db 946 FGALP--GGPGGPRPGNNGAINGALGSE--SSQGDMDNDGCGEDYDDVSEHLSNSLQ 1003
OY 766 EDEDEEED-----VYDEDSLAGSGSGEKA---ISVRGD 799
Db 1004 EGDRSRGDDFKSLRPEOKLRIDATGVVNTNPRRSASSHGHSGVSTAPSPVHAS 1063
OY 800 SE---EASGAEEEVG-----TVAAATAGKENDS---NEKTTQOSSLPFRP----- 839
Db 1064 SOVYIKSSSPARSEASOGALDLPRAAPTSSSSSRSLPRKEKRVSPSLRSPSGSSHAS 1123
OY 840 -----PPD-----SIDQ-- 846
Db 1124 ANILTSPLRPPTVGIDCLPKGLQNHNLQOQHNLQOQAALVAAAANQHNHMQMALDQ 1183
OY 847 -----PQMEQSSGVLCGKEE-----GKPER 869
Db 1184 EQLRREAERAOQRAAAAAAQAARQRTPRQARDQROEGGPGAGRPNPLMGARPPF 1243
OY 870 SSPASALTPREGATSVTLVEELSLQEAMRKEP-----GESSSRKACEVCGOAFPSQA 922
Db 1244 GMFPLNPLRPPTATQNMCMANMNOISQVMRAAPFNPLALSGVRS--TTCGICGYKTFPHS 1302
OY 923 ALEHQKTHPKEGRLFTCVFCRQGFLEERATLKKHMLAHNOVQ 965
Db 1303 ALEIHVRSHTKERP-FKCSICDRGFTTKGNLKQHMILT--HKIR 1342

RESULT 5
S42748
finger protein - fruit fly (Drosophila virilis) (fragment)
C/Species: Drosophila virilis
C/Date: 20-Oct-1994 #sequence_revision 26-May-1995 #text_change 24-Sep-1998
C/Accession: S42748
R:Schuh, R.
Submitted to the EMBL Data Library, November 1993
A/Reference number: S42748
A/Accession: S42748
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1402 <SCH>
A:Cross-references: EMBL:227444; NID:g426461; PID:g426462
C:Genetics:
```


Db 200 QPTNGLEKELEEQEPSEODHES- RRENSKTDKRGTEDRKA----- EREGQSMCD 252
OY 267 -----AGVGRSKHPTAPR----- PALPGSTDOLASPHLAPFTTGLLAQCICLARG 336
Db 253 ISSSLASSITTNDDPPAPNEPCLEMLQRTREYLD----- ASOSILHAQO- 299
OY 337 LEATASPGLLKPKNGSGELSGEYVPLEKPGGRHKRCFCAKVFSGDSALQIH- 396
Db 300 MGESEYASKEAOSREIF----- KIRCKYCGKIFGYSALQIHRSHTGE 346
OY 397 RPYKVCVGNRFTTGNLKVHFHREKYPHYQNMHPVPEHLDVYITSSG----- 447
Db 347 RPFVNCVCSKFTTKGNLKVYQHTQIFPPMLLPBGVAPN----- VGHSGQGOVGEQY 401
OY 448 -----LPGVMSVPEKAE-----EEAATPGGVKRPVLVASTTSLATESITLSTA 495
Db 402 PIRLPAPRYAPVGOEHQNOVEPEEIPVPOAEDLSKPVKEKE--KSHSPVECVKTPK 459
OY 496 GTATAPGLPAFNKFLMKAVEPKNADENTPPGS-----EGSAISGVASSTATL- 545
Db 460 EVKTDASLPSSSEK-----PEKEISKPVYTSRRNGSVKRRQTSVSPROGEDRENDIV 511
OY 546 --MOLSKMTSLPFWALLTHFKSTGSPFLPLCARALGAS-----PSETSKLQOLV 594
Db 512 EHMNTAKLVRSSA-----SRESOPAEYSIAQMERIDIKSWEDLIEIDKTSSTSKLQOLV 566
OY 595 EKIDROGAAVAVTSAAGAPTTSAPAPSSASSGPNQVYICLVLSCPALRLHYGONGE 654
Db 567 DNIENK-----LTDPRQCIFCQKVMSCRSSLOMIRHTTGE 602
OY 655 RPFCKVCGRASTRGNLRAHVGHKASPAARAQNSCPICOKFTNAVTLQOHVNMHL- 712
Db 603 RPFRCIKCRAFAFTKGNLKAHMSIHKIKPQRSQFCPCVHQKFSNGILLOQHIRHTMD 662
OY 713 ---GGO-IPNGGTALPEGGAQENG-----EGSYSGASSFPQ 748
Db 663 DGSGGQGVAAANGEBERUGIEDONSNSKSGTSPTLDFSTTIDSHSGQSESSQGDGDFDE 722
OY 749 QOSQO-----PSPEELSEEEDEDEEEDVTDSDS-----LAGRSES 788
Db 723 FMTMDSTDRSNSAATATPHPLERDRERERIRIPNOCSDERSHNPDLTGSRSES 782
OY 789 GGEKATSVAGDSEEA-----SGAEEVGT 812
Db 783 GEMPRAMDLSPPSSNSGRIFATGLANGATGGGSGNGGLPMLGMPMPNLLILMAAREEMHA 842
OY 813 VAAA-----AT----- 818
Db 843 LGHAHAKFPILLPFGPLGFMGLHPPRNVNCNLCFKMLPSLALESILQSEHAKEPATGHAOR 902
OY 819 -----AGKEMDSNEKTTQOSSLP-----PPPP--DSLD 845
Db 903 PHCSADGSPYGAKLTLNPLMFAKKPPSSSSSSSGEKLPESSNPPEAPMPATPIKEDPD 962
OY 846 QOPRP-EOGSGSVLGKKEGKPERSSPASALTPEGEATVTL-----VELSLQEA- 897
Db 963 QEOQLMVERBAS-----AGESGCGATSNYPQEDAGDAEOSLKMQLMHAHFPAPSLPFGQAL 1018
OY 898 MKRKEGESS-----SRKACEVCGOAFPSOALAEHQTHHKEGPLFTVCVFRQGLEEA 951
Db 1019 MSAGPPTSLDPPVNNKHKCHVCGRNRNFSSSSALQIIMKTHHGDKP-FOCNVCOKAFPTTG 1077
OY 952 TLKXHMLLAHNOVOPAPRGPONIAALSLVPG 983
Db 1078 NLKVHM-GTHMNTNPTSRGRRMSTLELPMRPG 1108
RESULT 7
20969
ypothetical protein F15C11.1 - Caenorhabditis elegans
Species: Caenorhabditis elegans
Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
Accession: T20969

R.Wilkinson, J.
submitted to the EMBL Data Library, April 1996
A:Reference number: Z19352
A:Accession: T20969
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-744 <Ntl>
A:Cross-references: EMBL:Z19352; PDB:1A95798.1; GSPDB:GN00019; CESP:F15C11.1
A:Experimental source: clone F15C11
C:Genetics:
A:Gene: CESP:F15C11.1
A:Map position: 1
A:Introns: 26/1; 75/3; 108/3; 145/3; 251/1; 453/2; 500/2; 613/1; 649/1
Query Match 7.48; Score 391.5; DB 2; Length 744;
Best Local Similarity 21.88; Pred. No. 1.1e-11;
Matches 194; Conservative 89; Mismatches 278; Indels 327; Gaps 32;
OY 219 ASPSELPGTGTASTKPLPLPSPKIPVQTSKTLASSSSSSSSSGAETPKOAFNHLVHP 278
Db 18 SKPRRMSGEDA-----MMSPL-----DLSTKSFDENNCEK----- 48
OY 279 LGSQHPSAGGVGRSHKPTAPSPALPGSTDOLASPHLAPFTT-----GLLAQCICLGA 333
Db 49 -----GAGG-----ALP-LEDRSNITLPHFSVPFANPQGFSLCAQIGNSS 87
OY 334 ARGLENTASFGLLKPKNGSGELSGEYVPLEKPGGRHKRCFCAKVFSGDSALQIH- LRS 392
Db 88 SRNVSTASTTSSCP-----IQCSOSFSSPAAITMVLDA 123
OY 393 HTGERP-YKCNVGNRPFTTGNLKVHFHREKYPHYQNMHPVPEHLDVYITSSGLPYG 451
Db 124 HDEQGLFSDVCTTTS-----NGQDIRH----- 149
OY 452 MSVPPEKAEEREAATPGGVKRPVLVASTTSLATESITLSTAGTATAPGLPAFNKFLV 511
Db 150 -----KCOKTLASRSTVSPSTIPSSCYFLS-----TPTPCL---QFSL 186
OY 512 MKAVEPKNADENTPGSEBSAISGVASSTATIMQSLKMTSLPFWALLTH- FKS 567
Db 187 NSIGTSEIREDEBEDMDVDEGEHVANQLFGLHLOKSDDKSKMS--LEFNAPPPFAA 243
OY 568 TGSFPLPCAR-----ALG-----ASPSETSKLQOLVIDROGAV 603
Db 244 FPNMPPFLMRQGFDPRAVDFAGRHNDNDDEALMELSTDEAKIRLV-----GDK 297
OY 604 AVTSAAGAPTTSAPAPSSASSGPNQVYICLVYLSCPALRLIYHGOGEKPFCKVCG 663
Db 298 AV-----PTTD-----PNOCTICRYVLSCKSALQIMHYRTHTGERPFCKICQ 339
OY 664 RAFSTRGNTLRAHVGHK-----AS 682
Db 340 RAFTTKGNLKTGHGVRKSKSPFGLPISLPQLAAMHONOHQIAPPRHINNPISAAS 399
OY 683 PAA-----RAONSCPICQKFTNAVTLQOHVNMHLG-----QIPNGSTA-----LP 724
Db 400 AAAAVALQIASOQCPIQGFNLNAGELAVHTEHRNSLTQPRVAPPTPTTIVQTFEVP 459
OY 725 EGGGAQENGSEST--VSGAGSPFQOQSQSPSEBELSEEEDEDEEEDVTDSDSLAG 783
Db 460 FETTPSLNATDSTQFNLANLISLAQLKNDSPNTDVSSVEEKTTRD-----DPRKMS 513
OY 784 RGSSEGGKATSVAGDSEASGAEEVGTAAATAGKEMDSNEKTTQOSSLPPPPDS 843
Db 514 LSPNSMSOSSSVRODILLESFEFEKL-----KLEPEPLLEOVSTTPNP----- 559
OY 844 LDQOPMEQSSGVLGKKEGKPERSSPASALTPEGEATVTLVEELSLQEAAMKREG 903
Db 560 -----KNEPDLAMQKMAATEPPRQMPV----- 585
OY 904 ESSSRKACEVCGOAFPSOALAEHQTHHKEGPLFTVCVFRQGLEKATLKKHMLAHNQ 963
Db 586 --LSKHQCGVCFKHFSSSSSALQIIMKTHHTGDKP-FKDCMGRAPFTTGNLKVHMGTHSWQ 642

964 VQPF-----APRGPONIALSLVPGCSF--SITSGLS 994
 643 QSPSRGRRIQFDVASSVTEKPMLOSPILPTSGAGASPLAMLGPNGLS 690
 RESULT 8
 742717
 DNA-binding protein Rc - mouse
 A:Alternate names: Ig kappa chain gene enhancer Recognition component
 C:Species: Mus musculus (house mouse)
 C:Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 18-Feb-2000
 C:Accession: T42717
 R:Wu, L.C.; Liu, Y.; Strandmann, J.; Mak, C.H.; Lee, B.; Li, Z.; Yu, C.Y.
 Genomics 35, 415-424, 1996
 A:Title: The mouse DNA binding protein Rc for the kappa B motif of transcription and for
 ew family of large transcriptional proteins.
 A:Reference number: 22238; MUID:97001141; PMID:8812474
 A:Accession: T42717
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-2282 <WU>
 A:Cross-references: EMBL:L46815; NID:91377885; PID:91377886; PIDN:AAB40884.1
 A:Experimental source: strain BALB/c; clone T1; thymocyte, brain
 C:Genetics:
 A:Gene: Rc
 C:Function:
 A:Description: binds V(D)J recombination signal sequence and kappa B motif
 C:Superfamily: HIV-Ep2 enhancer-binding protein
 C:Keywords: DNA recombination; transcription factor
 Query Match 6.6%; Score 346.5; DB 2; Length 2282;
 Best Local Similarity 23.0%; Pred. No. 5.2e-09;
 Matches 224; Conservative 114; Mismatches 336; Indels 299; Gaps 44;
 203 RQVLILGSGQYTGAPASPELPGTGTA---SSYKPLP---PLFSPKPVQTSKTLASS 255
 18 RRRLLTGEMAIQTIVSSSAP--YPSGTTAPSESATQELLATQPFSGP-----SQ 64
 256 SSSSSSSSGAETPK-QAFPHLYHPLGSOHPVSAGVGSRSHKPTPA-----PSPALPGST 308
 65 EKTGOQKRPARRPSIEASVNI-----SQLP-----QHPLTPAFMSPGKPHLLLEGST 111
 309 DQLI-----ASPHIAF-----PSTGTLAAQCL-GAARGLEATPSPGLTKKNG 351
 112 WQLVDPMRPEPSSFEVAPGSHQSQLPSHASILPPEELPGIKVTVPRPSQVSLKP--- 168
 352 SGEISYGEVWPLEKPGGRHKRCFCAYFGSDSALOIHLSHTGERPYKCNVCNRFETTR 411
 169 -AEAAHKKEKRP-QKP-GKYICQYCSRPCAKPSVLOKHIRSHTEGRRPYPCGPGCFSPKTK 225
 412 GNLKVHFRHREKYPHVOMDPVPEHLDVYTSSGLPYGMS---VPEKAEF----- 461
 226 SNL-----YKHKSHAHRIKAGLASGSSSEMYPPGLEMERIPGEFEETEGEST 275
 462 -----EATPGGCGVERKP-----LVASTALSAESTLLLSAGTA---TAPGL 503
 276 DSEETEGAAAGSPSTDVLPKPKHPLSLSSLYSSGSHGSSQERCSLSOSTPSPLEDPAPFA 335
 504 PAFNKFEVL-----MKAVEPKNKADENT--PPGSEGAISGVAESSTATLM 546
 336 EASSEHPLSHKPEDTHTIKQALALRSEKKKLEIEQTFLSPGSGTSESVFSSESSEAEQ 395
 547 QLSKLTSLSPSMALL-----TNHFKSTGSRP-LPLCAR---ALGASSESKL 590
 396 QVSPPTNINASTYAEIIEGKCGRIGRTSMALASTSTQPLPLPSSEDKPSLVLVSPRTQVI 455
 591 QQAVEKIDROGAVAVTS-----AASGAPTSAPAPSSS----- 623
 456 EHTTKLTITNEAVVDTSDISDVKKPRRSILTRSSSVSPKSSLYRDSLSHCEKTKQESL 515
 624 -----ASSGPNOCVICLRVSLCPRALRLHYGQHGGERPKCKVCGRATSTRGNLRAHF 676

516 ISLQHPSPSTHP---VPLRSHSMPSA-----ACTISTHHTHTERSY--SP 556
 677 VHKKSPARAQNSCPIQCKKFTNAVTLQCHVM---HLGGIIPNGSTLLPREGGAQEN 733
 557 DDHVADPEVPSRNT-PV-----FTS-----HPRMKRIHAATLEPLGG----- 592
 734 GSEQSTVSGASFPQOOSQPS-P-EELISEEEDEEEDVDTD-----ED 779
 593 -----EVSSEBPGPSSKDPSTSKPDSDEPEKESDLTKTKKGFKTKGANYEC 638
 780 SLARG------SESGEKAISVRGDSSEASGAEEVGYVAAATAGK--E 822
 639 TICGARYKKRDNEYANKKYCELOITKAHSVGAHEVEKTAQEPESQSMHMYKLATLE 698
 823 MDSNEKTQOSSLPPRPPPSLDQPPMEQSSSVYLGGKEBEGKPRSSPASALPEGE 882
 699 LTPLRRRRKSSLSGDEEPPAFACPGSETJHNRPLGSTK---SPKASASASL-EDPR 754
 883 ATSVTLVEELSLQEAMKKEPESSSRKACEVCGAFAPSQAALEEHQKTHPKEGPLFTCVF 942
 755 ASSPGLPSQ-----ELGQNGGRG-EGCPKFKTVIQTSSFEKSDPEQP----- 798
 943 CRQGFLEATLTKHMLAHNOVPPAPRG-----PONTIALSLVPGCSPTS 990
 799 -----SGLEEDKPPAOFSSPPAPRGSAHSLOPRLVRPNIOVPELIVTEEDPRD 850
 991 TGLSPPRKDDPT 1003
 851 TEPEPPEKEPEKT 863
 RESULT 9
 B60191
 Transcription regulatory protein Evi-1, short form - human
 N:Alternate names: ecotrophic viral integration site 1; oncogene Evi-1
 C:Species: Homo sapiens (man)
 C:Date: 03-Mar-1993 #sequence_revision 03-Mar-1993 #text_change 29-Aug-1997
 R:Morishita, K.; Parganas, E.; Douglass, E.C.; Ihle, J.N.
 Oncogene 5, 963-971, 1990
 A:Title: Unique expression of the human Evi-1 gene in an endometrial carcinoma cell 1
 A:Reference number: A60191; MUID:90326419; PMID:2115646
 A:Accession: B60191
 A:Status: not compared with conceptual translation
 A:Molecule type: mRNA
 A:Residues: 1-727 <MOR>
 C:Genetics:
 A:Gene: GDB:EV11
 A:Cross-references: GDB:119889; OMIM:165215
 A:Map position: 3q26-3q26
 C:Keywords: alternative splicing; DNA binding; zinc finger
 Query Match 6.3%; Score 330.5; DB 2; Length 727;
 Best Local Similarity 24.2%; Pred. No. 8.9e-09;
 Matches 150; Conservative 82; Mismatches 231; Indels 157; Gaps 30;
 368 GGRHKRCFCAVFGSDSALOIHLSR-HTEGPRYKCNVCGRFTTRGMLKVFHRRREKYP 426
 128 GKHYECENCAVFTDPSNLQRIHRSQHVCAKAAACBEGCTPATSSGLKQKHHSVYP 187
 427 HVQMPN--HVPEN-----LDVYTSSG---LPGMSVVP-----EKAEEAATPGGG 469
 188 FISRQSMYPPDRDLRLSLPLKMEPQSPGVKTKQKGSSSPEDLTTRKRDKEKPLP--- 244
 470 VERPPLVASTALSTESTILSTASGTATAPDLPAFNKNVIMKAVAP-----KKKA 521
 245 VPSKPPVYPATSDQDPLDSLMSGRASRASGT-----KLTERKKNHVFGGKGS 291
 522 DENTPPGSEGAISGVAESSTAT-----LMQSLKLTSLSPSMALLTNHFK-STG----- 569
 292 NVESRPASDGS-----LQIARPTPFMDPIYRYVEKRLTLDPLEALKKKYLPSPGFLFHP 346
 570 SFPLPLCARLGAASPESTSKLOQLVEKIDROGAV-----AVTSNAS-GAPTTSAFA 619

hypothetical protein K06A9.1a - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 18-Feb-2000

C:Accession: T34434

R:Geisel, C., Gattung, S.

submitted to the EMBL Data Library, December 1996

A:Description: The sequence of C. elegans cosmid K06A9.

A:Reference number: 221525

A:Accession: T34434

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1,2232 <GET>

A:Cross-References: EMBL:U80846; PIDN:AMC70890.1; GSPDB:GND0028; CESP:K06A9.1a

A:Experimental source: strain Bristol N2; clone K06A9

C:Genetics:

A:Gene: CESP:K06A9.1a

A:Map position: X

A:Introns: 38/1, 75/3, 103/3, 132/2, 158/2, 222/1, 1088/1, 1367/1, 2039/1, 2049/1, 2075/1

Query Match 6.0%; Score 316.5; DB 2; Length 2232.

Best Local Similarity 21.2%; Pred. No. 1.4e-07;

Matches 231; Conservative 188; Mismatches 435; Indels 295; Gaps 48;

QY 51 ONACGTDPPVWVYIIGGQENPNNS-----ASSEPRPGHNHPQYMDTHSNPPDGGSSV 104

Db 498 QASASSPEPTMSTVSG---PTGSTVYVPGSSTSPADSSPPPS-----SSPASTGSTI 548

QY 105 PTDPTWGPERRGEE-----SCHFLVAATGTAAGGGGLILASPILGATP-----LPPE 153

Db 549 TIS-----GSSSIIVTSGSTVSGSTGTSQS-----TLASSP--ATPGSSSTVPS 593

QY 154 STPAAPPPPPPP-----PPPGVSGHLNPLILEELRYLQRIQHMO--- 196

Db 594 SSPQSSQSPAPMTGSTPTSPQTSQSSPSPSPNPS--STPTGSSQSTTPEGSTASSPTGS 652

QY 197 -----MTQICROVLLG-SLG-QTVGAPAPSPSEL-----PGTGASS 232

Db 653 TGSTPSVAEVTYQSIVTSGSSSLCTGTSNPSPSLSPTSGMSTLTSEPSSTQSSG 712

QY 223 TKPLPLPFIKIVQTSKTLASSSSSSSSSGAETPKQAFPHLYHPLGSOHPPSAGVGR 292

Db 713 AOSTLTTPSP-NPSQSTSLSSSTSGATTSGSGAGTWT-----SPQSSSVGS 760

QY 293 SHKPPAPAPALPGLSDQLLASPHLAPSTGLLAQCGLAAGLEATASPG--LLKPKN 350

Db 761 SGGST---SPAASTSGEMTSGSTQTPGSSVSTSAIITLTSQOVSVTSNGSTVTRPST 817

QY 351 GSGELSYGE--VWGPLEKPCGRHRCFCAKVFGSDALQIHLRSHTGERPKYKNCVGNRF 408

Db 818 VGSSTSSGSTVTVGSTM-----ASTSGSSVA----- 843

QY 409 TTRGNLKVHFHRRREKPPHYQMNPHVPPEHLDYVTTSSGLPYGMSVPEKAEBAATPGC 468

Db 844 -----SSPA PSTSQNPNPSTSSGSSMTQSPFYQPOSTAP---VASTTSP 887

QY 469 GVERKPLVASTATLSATSTLLTSLTSAGTATAPGAPAFKPFVLKAVEPKNKADENPPG 528

Db 888 GSPGTLT--TSTSPSPSQSTTIGSTG--STSPGISTTS-----EEMTSGGSTPTPG 935

QY 529 SEGSAI--SGVAES-----STATLMQSLKMLTSLPSWALLTNHFKSTGSPFLPCARALG 581

Db 936 STGSTVTPSTVSDSTSGSGSTVTVGSTRGSSSPIDSTQNTNPSTSSGS----- 984

QY 582 ASPESEFSKLOQVLEKIDROGAVVATSAASGAP---TTAPAPSSASSSGPNQCIVLRY 637

Db 985 SMTSTQPOSSQSTSPVSSSTGATSS--SGSPGTLTSTISPSPSPSTIGSSQGSTSPV 1042

QY 638 LSCPALRLHYGCHGERPFCKKVCGRAFS---TRGNIRAFV-----GHRASP 683

Db 1043 STISQGSTERTPGSTGSTVTRKSTVSGSASSGSTATMGSTEASTSGSSSTSPNPSQSTSP 1102

QY 684 AARQNSCPIQCKKFTNAVTVLQIHVRHMLG-----OIP--NG 719

Df Db STSGATSSPSSGCTTLTSTSPSQSSTIGSSQGSSTSPVSVTTSGDMISOGSTQLPGSTG 1162

Oy 720 GTALPEGCGAOMENSGESTVSAGSFPOOQSOPSPPEELSEEEDEEEDVDDBD 779

Df Db STVQPSTGSGSTSTSGELITTSQGTOFP-RSSLSTSPAISTSTQOSVTNSBPSVTVOFS 1221

Oy 780 SLAGRGSESGETKKAIVRDDSEASGAEEEVGYAAATAACKEMDSNKTTOOSSLPFPP 839

Df Db TV--RGSTISSG-----SYTTGTGSTESSSTSGLSSATSLS-SSP 1257

Oy 840 PPDSLDPQMBOGGSSGVLGKEEGGRPERSSSPASALTPEGEATSVTLVEELSLOEANR 899

Df Db VPTSGSPRPSTSGSS-----TPTRPMPSOSTSPYVS-TTGEMTSHGSTOPTSTGSTV 1310

Oy 900 KEP---GSSSRKACEV-----CGQA---PQAALEHOKTHREGPLFTVCFCMQ 945

Df Db TQPSVTVSGNSSGSTVITJGSSSEASTSGSFRTPSSSIS----PVPRSSPIPTTFASS 1364

Oy 946 GFLERATLKHKMLLAHQVPAPHGPNIALSLVPCSPSTST-----GLSPFR 998

Df Db TSSTIDVSVSTSTSLAPLSSLP-----STVBSSTGSSFSTSEGSSKASSSPVPS 1416

Oy 999 K--DDPTIP 1005

Df Db QTSSTPTNP 1425

RESULT 12
T13893
gene hindsight protein - fruit fly (*Drosophila melanogaster*)
C:Species: Drosophila melanogaster
C>Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 17-Nov-2000
C:Accession: T13893
R:Tip, M.L.R.; Lamka, M.L.; Lipshtz, H.D.
Development 124, 2129-2141, 1997
A>Title: Control of germ-band retraction in *Drosophila* by the zinc-finger protein LIN
A:Reference number: Z17807; MID:97330681; PMID:9187140
A:Accession: T13893
A>Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: mRNA
A:Residues: 1-1920 <YIP>
A:Cross-references: EMBL:U86010; NID:g2769709; PID:g2769710; PIDN:AAB95640.1
C:Genetics:
A:Gene: hindsight
A:Cross-references: FlyBase:FBgn0003053
C:Function:
A:Description: probably function as a transcription factor
C:Keywords: nucleus; zinc finger

Query Match 5.9%; Score 313; DB 2; Length 1920;
Best Local Similarity 20.0%; Pred. No. 1.7e+07;
Matches 249; Conservative 137; Mismatches 438; Indels 424; Gaps 56;

Oy 16 GEPAETGGASEEDHPOVCAKKCAOFDTPEFLAHONACSTDPPVMWVIIGGENPNSS- 74
Db 382 GWTLASGAESYESD-----ASCSS--TDVSSGHSHRSRSSS-----LNNNNNNH 423

Oy 75 -----ASSEPRREGHNPNQVMDTEHSNPPSGSVPTDPTWGPERGERGESG 121
Db 424 KANNNLKDELFELEVSTEDDOTTENKORRKTTINNIIIESEQEDMD-----DEADDAV 478

Oy 122 HPFLVAATGTAAAGGGGILLASPKLAGATPLPESTRPAPP PPPPPP-----PGVSG-- 173
Db 479 AMLTSPDVAT-----LLAGASASGA-----ASRSPHPSASPALLISCAPGASDFE 527

Oy 174 -----HLNIPLILELRVL-----QQRIHQNMOTEOICR-QYVLGSLGQTVG 216
Db 528 TLPALCVHLDAHMSDIKAKCDCEVFATHROLOSH-----CCRLPNALAGLPLPLIG 580

Oy 217 APASP--SELPGNTASTAFKPDLPLPFPIKPQTKTLASSSSSSSSSGAEFKQAFFH 274
Db 581 ASSSPLHNEEPDEBEHDDEUL-----EQKERLASGED----- 614

QY	688	-----QNSCPICQ-----KKTNNVNTLOOHNR-----MHLGGQ	715
Db	1133	HGHGHGSHGHSHAPISEQVYKCALLAQOLKAHKNTDLOOLAHGSSVAGNPLHLGYR	1192
QY	716	IPNGGTALPEGGAOENGSEQSTVSGAGS-F-----PQOOSQOPPEEELISEEEDEDEE	771
Db	1193	LITNPS-----PMHNGSSQGGQCATALDDDEPKLITDEBENHDEHVEDVDVDFEEDDEE	1249
QY	772	EEDVTDSDSL-----AGRGSESGE--KAISVRGDSEBAGAEVEGVAAATAAGKM	823
Db	1250	MDEPEDEPELIDQPAKKEAEEQELRPLPOLGTCKEAAQKMAETI--LEQAIIKAKPL	1307
QY	824	DSNKRTTQOSSLPLPPRPDSLDLP---QPMEGSS--GVUGKEEBCG--PESSSP---	873
Db	1308	-SPPTKEMASPPANVTVAITTMQEPATITAPSTNPSLKTMAIAQAEVYGKSLKEVASPEFKD	1366
QY	874	-ASALTPEGE-----ATSYT-----LVEELSLDE-----	896
Db	1367	ESQDLVPAKLVLDNATSQMGNFNSYFRPSDVANHMQSDDEGLVAGGSNASESNSSTEDV	1426
QY	897	---AMRKPEGSS-----SRKACEVCGAFAFQQAALBEHQKTHPKRGPLFTCVCFRGAF	947
Db	1427	TSSSSSSPPKKKSAVSLAPNRVFSPCQGMFPWSSSLRHHILHTHGQK-PFKSCHPLLF	1485
QY	948	LERATLKKHMLLHHQVQ	965
Db	1486	TTKSCNDRHLLRKHGAVE	1503

RESULT 14
I50620
prockr2 - chicken (fragment)
C:Species: Gallus gallus (chicken)
C:Dates: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 21-Jul-2000
C:Accession: I50620
R:Schutz, B.; Nlesing, J.
Gene 148, 227-236, 1994
A:Title: Cloning and structure of a chicken zinc finger cDNA: restricted expression in d
A:Reference number: I50620; MUID:95047430; PMID:7958949

A:Cross-References: EMBL:X56805; NID:g577018; PIDN:CAA40140.1; PID:g577019
C:Genetics:
A:Gene: CK12

Query Match	5.88;	Score 307;	DB 2;	Length 1173;
Best Local Similarity	21.48;	Pred. No. 2e-07;		
Matches 189; Conservative	90;	Mismatches 314;	Indels 293;	Gaps 36;

[illegible]

Db	266	РУКСОСРСРSPG-----	QLAGLLGNORG-----	HSATEPHRPAPTPPTSPSPSERPY	312
OY	630	OCVLC-----		LRVLSCPR	642
Db	313	QCTEGKAFKSSGLRYMRDHTGERPYKSCSPCAFKKSSLLAHQRYNTGLRAYKCS			372
OY	643	A-----	LRLYGQ-----	HGERPEKCYCGRA	665
Db	373	CGLPEKWSHLQYHLRLHTGERPYRCPDCKAFKNTSCGLPHQDLHTHTGERPHACIPCKGA			432
OY	666	FSTRGNLEAH---FVG-----		HASPAARQONSCPIQCKETNA	701
Db	433	FQTSYMLRKHONHTHTGERPYACSHCKFTTTHSSNLQHLHORTISSARPHOCPLDCKAFVMA			492
OY	702	VTLOQHYVNMHLGQPIRNGSTALPEGGGAQOENG-----	SEOSTVSGAGSPPOQOSQO-		753
Db	493	SYLQNLHTNHAAG--PKGS--	PRALPTPQORGVLAALSLLEVTPARDAHTFLLQTPQG		547
OY	754	-----PSPRE--ELSEEEDEDEEEDVTDDESL--	AGRGs-----		786
Db	548	LQLPSPRPAPRKLILPTAPRPPRHQOESPSPGOSLLVPSTGTLPTPLRLQAVTAYP			607
OY	787	ESGGEKAISVRDSEB---ASGAEEBVTVAATAATGKEMDSNEKT--TOOSSLP	PRPPRP		842
Db	608	QCTGIGLVLQGLRPPQRLPHRAGIRHQGAVEGAAY--	RLQANEVYNVQLOALRPPDVT		664
OY	843	SID-----QOPMEDGSSGVLGKGEKGKPEKSSSPASALYTBEGEATSVTLVELS			893
Db	665	SIPLOATMDTNYQLOALPQ-----	PTDVTNYQLOQTEVTNYO---		701
OY	894	LOEAMRKKEGESSSRKACEVCGOAFRPSQALBENOKTHPEKGRPLTCVFCROGFLERATL			953
Db	702	LQALLQRPDVTNYQIQMATEVANYQL--QALSQRPDVTNIE-----	LQAAEV		745
OY	954	KXNHLIHHQVOPRAPHQONTAALSLVBCSPSTIGSLPPRP			998
Db	746	TN-----AHQOLVR-----QSEVTNYQLOOG--	EVVNYQLOQLTPQ		779

RESULT 15

finger protein 2, placental - human
C:Species: Homo sapiens (man)
C:Date: 20-Oct-1989 #sequence_revision 20-Oct-1989 #text_change 01-Dec-2000
C:Accession: B32891
R:Bellefroid, E.J.; Lecocq, P.J.; Benhida, A.; Poncelet, D.A.; Belayew, A.; Marfai, DNA 8 377-387, 1989
A>Title: The human genome contains hundreds of genes coding for finger proteins of th
A:Reference number: A32891; MUID:89377476; PMID:2505992
A:Accession: B32891
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-651 <BRL>
A:Cross-references: GB:M27878
C:Superfamily: zinc finger protein ZFP-36; LIM metal-binding repeat homology
C:Keywords: tandem repeat

[illegible]


```

QY      606 TSAASGAPTTSAAPRSPSSASSSGPNOCVTCILRVLSCLPRAALNLTQHOGGERPPKCVCGSRA    665
           : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      304 CGKAFTOSTSLKHNRLLVNAOHNEFPYRCOECGVRRHRPRRLMLHRYHNHTGEYPYCKRECEPS    363
QY      666 FSTGNLNAH-FVGHKASPAAARNONSCPTIOKKFNTNVTLTQQHYRMHGGOIPNG---GT       721
           + | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db      364 FLRLRLLEVHLVH-----AGROPNRCPSCGAAPRSSLRLEHNHCANAAAQAARPREECT    419
QY      722 ALPEGGAAQENGSEDSITVSGAGSFPOOQSOPRSPEELSEEEEEDEEEDVDSDSL          781
           : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      420 CGKKGVAARLQAH-EAHAAGAEGVELAKEPPRAPRARTR-----APVASP         466
QY      782 AGRGSESGEKALISVAGDBDEASGAEEVEVTVAATAAGAKMDSNEKTQ-----QSSLP     836
           : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      467 AALGSTATASPAPARARRGLECSEC-----KKLFSTETSLQVHRRHTHTGER   512
QY      837 PPPPDSLDOQPMIEGSSGVLGKKEGCKPERSSSPASALTPEGATSVLVEELSLOE    896
           : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      513 PYPCCDC-----GKAFRGRT-----HLKD        531
QY      897 AMRKPESSSRKACVCCQAFPSQAALAEHQKTHPKERPLFTVCFCQGFLEPATLKXH    956
           : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      532 HRRLLTGERRPF--ACEVCGKAFAISMRLAEHRIHTGERP-YSCPDCKSKYSFSFMKMH    588
QY      957 MILANHVOQ    965
           |||
Db      589 R--KTHOOQ    595

RESULT      18
S70619
finger protein bowel - fruit fly (Drosophila melanogaster)
C:Species: Drosophila melanogaster
C:Date: 19-Mar-1998 #sequence_reviation 17-Apr-1998 #text_change 17-Nov-2000
C:Accession: S70619
R:Mang, L.; Coulter, D.E.
EMBO J. 15, 3182-3196, 1996
A>Title: bowel, an odd-skipped homolog, functions in the terminal pathway during Drosophila embryogenesis
A:Reference number: S70619; MUID:96272178; PMID:8670819
A:Accession: S70619
A:Molecule type: mRNA
A:Residues: 1-744 <MAN>
A:Cross-references: EMBL:U58282; NID:g1388165; PIDN:AAB17949.1; PID:g1388166
C:Genetics:
A:Gene: bowl
A:References: FlyBase:FBgn0004893
A:Map position: 2L
C:Keywords: transcription regulation; zinc finger

Query Match              5.6%; Score 293.5; DB 2; Length 744;
Best Local Similarity    17.2%; Pred. No. 5.4e-07;
Matches 174; Conservative 88; Mismatches 327; Indels 425; Gaps 24;

QY      90 MDTEHSNPDDSGSVPTDPTWGPERRGEESSGHFLVAATGTAAGGGGGLILASPKLG--    146
           ||| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      1 MPTESSSSSEISGGGAIPLMLRPSRMDQFMNS----MAAAAAAVGGGGLPGADRNGSGS    56
QY      147 -----ATPLRPESLP-----APPPPPP-----            164
           : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      57 GSDGGSONGNGDSRNSSASRSISAYETOLAYOOHLAGLGMRRPPRPPSHREISAIVPLP    116
QY      165 --PPPPGVSGHNLIRPLELETRYLQQRQIHQMOMTEQICROVLILLGSIGQTGAPASPS    222
           || : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      117 TGYVRPGSNSENSYEIIAMADKKELALRE-----AAAAAAMLGRGFGPGPG    164
QY      223 ELPGCTGAASSTKPLLPLFSPIKPVQTSKLASSSSSSSSSSSGAETPKQAFPHLYNPLGSO    282
           : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      165 -VPPPGV-----PGSPYGAGS-----SP---PGCAAAAAALFPPLGLGROMHAGLD    178
QY      283 HPPSAGGVGSHKRTTPAPSPALPGSTDQLIASPHLAFTSTGILLAOCTGAARELENTAS    342
           : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      179 PPPLYLG-----PGSPYGAGS-----SP---PGCAAAAAALFPPLGLGROMHAGLD    221

```

Query Match 5.6% Score 293.5; DB 2; Length 1191;
Best Local Similarity 22.5%; Pred. No. 8.9e-07;

Db 343 POLLRKKNSSGELSYGEWMPLERLECKRGRRHKRCRCACVWFGSDSALQIHNRSTGEPRKYCN 402
||| : : : : : ||| : : : : : ||| : : : : :
222 RRLR-----APGRASRPKKOFTCKFCNROFTSYNLLIHERTHDERPYSD 269
||| : : : : : ||| : : : : : ||| : : : : :
OY 403 VCGNREFTTGNLKVHFNHREREKYPHVQMNPHRVENLDVVITSSGLPYGMVSVPKEAEE 462
||| : : : : : ||| : : : : : ||| : : : : :
Db 270 ICGKAFRRDCHLRHRYIHSKKP----- 293
||| : : : : : ||| : : : : : ||| : : : : :
OY 463 AATPGGVERKPDLVAITTALESLLTLSTAGTATAPGLPATNKFYLMKAVERPNKD 522
||| : : : : : ||| : : : : : ||| : : : : :
Db 294 -----FK----- 295
||| : : : : : ||| : : : : : ||| : : : : :
OY 583 SPSETSKLOQLVEKIDROGAVALTSASGAPTTSAPESSASSGPNOCVICLVLSCP 642
||| : : : : : ||| : : : : : ||| : : : : :
Db 296 -----CTECGKGFCOSR 307
||| : : : : : ||| : : : : : ||| : : : : :
OY 643 ALHLHYOHGGERPPPKCYCGRAFSTRGNLRANFVGHKASPAAQNOSCPIQOKFTNAV 702
||| : : : : : ||| : : : : : ||| : : : : :
Db 308 TLAVNHKLIMEESPRKCVCPSRSFNQRSNLKTILLT---TDHKRYECSGCKVFRRNC 363
||| : : : : : ||| : : : : : ||| : : : : :
OY 703 TLOOHVRNHLGOLINGTA-----LPREGGAA 730
||| : : : : : ||| : : : : : ||| : : : : :
Db 364 DLRNHALTHAVGVNSGDVDVNGEDEARNLSDGEDSLLEVDSFRQSPVHNLGSGSG 423
||| : : : : : ||| : : : : : ||| : : : : :
OY 731 QENGSEOSTVSGAGSFPOOOSOOP---SPEEDISE-----EEEDDEEEDVTDE 778
||| : : : : : ||| : : : : : ||| : : : : :
Db 424 EKSESERMLRKKAIDHEESEEFDDFDDEEELODLPRVNDLPREEDDFPREDEQAE 483
||| : : : : : ||| : : : : : ||| : : : : :
OY 779 DSLAGR-----GSSEGKATSVRGDSEASGAEEEGTVAAATAAGKEMDSNEKTTOQ 832
||| : : : : : ||| : : : : : ||| : : : : :
Db 484 VALARFOASKAAASQSSTSVGTKEPRQGVTHGHNEGGETTYMTRHGKHODERPNCSI 543
||| : : : : : ||| : : : : : ||| : : : : :
OY 833 SSLPPP-----PPSDLDPQPMEGSSGVLGKEBGRKERSSPA--SALTREGEA 883
||| : : : : : ||| : : : : : ||| : : : : :
Db 544 ASLVPREFVRSVPRGAGGRP-----APGARPTIONRPNHILPRMGD- 590
||| : : : : : ||| : : : : : ||| : : : : :
OY 884 TSVTLVEELSIOEARKEEGESSRKACEYCQAARPSAALDENQKTHKEGFLTCVFC 943
||| : : : : : ||| : : : : : ||| : : : : :
Db 591 ----PYLPLTHVRDDLHHKSNTLSKAGVPRPHPTPTTIITQPESGKPLNDLHSPHEA 644
||| : : : : : ||| : : : : : ||| : : : : :
OY 944 RGFLERATLKKNMLLNHOVOPARPONIAALSLVPGCSPTSISTGSLSP 997
||| : : : : : ||| : : : : : ||| : : : : :
Db 645 MPSELGISIPMKRIIPA-PTLDLMDPHNNHGLGQTGYD--SPSIYALNNSRNP 695
||| : : : : : ||| : : : : : ||| : : : : :
RESULT 19
S35305
zinc finger protein ZNF91 - human
C/Species: Homo sapiens (man)
C/Date: 31-Dec-1993 #sequence_revision 02-Jun-1994 #text_change 05-Nov-1999
C/Accession: S35305
R:Jellifroid, E.J.; Marine, J.C.; Ried, T.; Lecocq, P.J.; Riviere, M.; Amemiya, C.; P
EMBO J. 12, 1363-1374, 1993
A>Title: Clustered organisation of homologous KRAZ zinc-finger genes with enhanced ex
A:Reference number: S35305; MUID:93223677; PMID:8467795
A:Accession: S35305
A:Molecule type: mRNA
A:Residues: 1-1191 <BEL>
A:Cross-references: EMBL:L1672; NID:g186773; PIDN:AAA59469.1; PID:g186774
A>Note: the authors translated the codon GCA for residue 750 as Thr and GCT for resid
C:Genetics:
A:Gene: GDB:ZNF91, HPE7, HTF10
A:Cross-references: GDB:132284
A:Map position: 19p12-19p12
C:Keywords: DNA binding; zinc finger

Matches 141; Conservative 75; Mismatches 239; Indels 173; Gaps 24;	
OY 365	ERKGRHKRCPCAVFPGSDSALQIHLRSHTGERPKVCNCGNFTTRGNLKVHHRREK 424
DB 627	EKR---YKCECGGAFSHSSALAKHRIHTGKPKYCKRCGKAFSSNSTLANHITTEE 683
OY 425	YPHVQNNPVRPHLDVYTTSSGLPYGMSVPEPKAEFEAAATPGGVERKPLVASTALSA 484
DB 684	KPKCKECEDTKFKRLSTLTKHKIIHAGEKL--YKCEE-----CGKAFNR 725
OY 485	TESLTLSTASGATATAGLPANFKFVLMKAVPKKADENTPPSGSAGSAISGVAESSTAT 544
DB 726	SSMLT-----HKFI--HTGKPKYKCEE-----CGKAFNMSSS 756
OY 545	LMQLSKLMTSLPS-----W-ALLTNHFK-STGSFPLPL--CARALGASPTES- 588
DB 757	LTKHKRIHTREKPKCKECCGKAFWSSSTLTKHKRIHTGKPKCECGKAFSSSTLTKH 816
OY 589	KLQQLVEKIDRGAAVAVTSAAAGAPTTSPAPSSSASSGPNOCVLCRLVLSCPRALRLAY 648
DB 817	KTIHTGEK-----PYCKECCGKAFKHSSALAKHK 845
OY 649	GQHGGERPKCKVCGRAFTSRGNLRANFVGHKASPAARQNSCPTCQKFTNAVTLQOHV 708
DB 846	IIHAGKLYKCECGKAFNOSNSTLTKHIHTGKPKSSE---CDKAFWSSSTLTKH 901
OY 709	RNHLGQIPLNGGTALPEGGAAQENGSEOSTVSGASFPOQSQOPSPREELSEEEED 768
DB 902	RIH-----TRBKPYKCECGKAFS-----QPSHLTKHKRMHTGKPKYC 940
OY 769	EEDEEDVDEDSLAGRSESGEKEAIVSGDSEASGAEEVGVYAAATAGKEMDSNE- 827
DB 941	EECGKAFSSSTLTKHIHTGKPKYK---CECGKAFRKSSTL---TEHKIHTGK 992
OY 828	-----KTQOSSLPPLPPPPPSLDLPQPMEOGSSGYLGKKEGKPRRSPASALTP 879
DB 993	PYKCECGKAFSSS-----TLTKHTRMHTGE-----KPYKCECGKAFNR 1033
OY 880	EGEANSVTLVEELSLQEMARKPEGESSSRACVCGQAFPSQALAEHOKTHPEKPLEFT 939
DB 1034	SKSLTKTHKII-----HTGKPKYK--CECGKAFISSTLNGHKRIHTREKPYK 1079
OY 940	CVFCROGFLEATLTKKMLLAHHQOVPF 967
DB 1080	CEECGKAFSSSTLTKHKRL--HTGKPKY 1106
RESULT 20	
143676	
hunchback-related protein - Caenorhabditis elegans	
Species: Caenorhabditis elegans	
Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 21-Jul-2000	
Accession: 143676	
Fay, D.S.; Stanley, H.M.; Han, M.; Wood, W.B.	
Rev. Biol. 205, 240-253, 1999	
Title: A Caenorhabditis elegans homologue of hunchback is required for late stages of	
Reference number: 222622; MUID:99117349; PMID:9917360	
Accession: 143676	
Status: preliminary; translated from GB/EMBL/DBJ	
Molecule type: mRNA	
Residues: 1-982 <FAV>	
Cross-references: EMBL:AF097737; NID:94323034; P1DN:AAD16170.1; PID:94323035	
Gene: hbl-1	
Query Match	
Best local Similarity 5.5%; Score 292.5; DB 2; Length 982;	
Matches 214; Conservative 134; Mismatches 394; Indels 499; Gaps 46;	
OY 5	SERSRGLG-----VPAGE-PAELG--GDAS-----EEDHPVCACCAQFTDPT 45
DB 49	SANNSLGDMMKCCVPGSEIPQHLGPFNGSSLTMLTAQDPGEKIHDP-----GGVSPK 102

OY 46	E-----FLAHONACSTDPRVMYIIGGOENPNSSASSEPREGKNNQVDMTEHSNPPDSC 101
DB 103	EDGRKSEHNYSYD-----VSASQSPSNDGAQSDSTSEDIHIDECAMETEMDTEKD 154
OY 102	SSVPTDPTWGERGESESSGHFLVAATGTAAGGGGLIASPAGIATPLPESTPAP--- 158
DB 155	SLTKPEQATPKLEGSQSKSESTVEGTS-----SNVYTSBPVQMPQPIVYIP 205
OY 159	-----PPPPPPPPPPVGGSHNLPILLEELRYLQORQHOMQMTQICROYLLGSL 211
DB 206	SFLKNSLPAPRIPPTTOSANVERNSNSPIEFALLT---TL5QOQFAEVAFAAKIRKS 261
OY 212	GQTVGAPASPELTCGTASSTKPLPLPSPIKRPVQTSKTIASSSSSSSSSGAEFTKQA 271
DB 262	SESIGFORS-----GTSA-----FLNTEPEMSSSANNNDDEA----- 295
OY 272	FPHLYHPLGSHQHPSPAGGVGRHKPTPAPSPALPGSTDLIASPHLAFPTTGLAAQCL 331
DB 296	-----RPPGL-----GPVALPPTQNGQTPMLVCPICGFMCPSKFH 350
OY 332	GAARGLEATASPGILAKRNGSGELSYGEVMGPLEKPGGRH-----KRCPCAVFGSDSA 385
DB 316	-----RPPGL-----GPVALPPTQNGQTPMLVCPICGFMCPSKFH 350
OY 386	LQIHLRSHTGERPKVCNCGNFTTRGNLKVHHRREKYPHVQNNPVRPHLDVYITS 445
DB 351	FNSHMNTH--GD--HCCSMCDVTSRTEGRLLKHM---RES-----HYEEDQIRAGFES 396
OY 446	SGLPYKSNVPEKAEEBAATPGGVERKPLVASTALATESITLTLSTAGTATAGLPA 505
DB 397	-----EPAKESASSP-----KNLST--SKDGSATSPINEI 424
OY 506	FNKFVLMKAY--EPKKNKADENTPPSGSAGISGVAAESSATLMQLSKMTSLSPWALLTNH 564
DB 425	FNLSITMASLIDSTNN-----AVSISTTEQPSAL-----SALTLD 460
OY 565	FKSTGSFPLPCARALGASBSETSRLQOLVEKIDRGAAVAVTSA---ASGAPTTSPAPS 621
DB 461	MSTPFLSTLTAHSPGVASLDQIKATISENPSPMPEGGINLASLGVNAITGDDPSPE 520
OY 622	SS-----ASSSP-----NQC-----VLCRLVLSCPRA----- 643
DB 521	KQSNCGRRSSSGKIKIFKCKQCGHQSLSKDOWANAHARTHIAPAEQLNCOHCNFTVEYKH 580
OY 644	LRINHGGERPKCKVCGRAFTSRGNLRANFVGHKASPAARQNSCPTCQKFTNAV 702
DB 581	HLEIYHNRNIIGSKRPFQCKKCAIYNKVNSKMLNSHKSH---TNHQFRMCDCTYATKYCH 636
OY 703	TLQOHVNMHLGQIPLNGGTALPEGGAAQENGSEOSTVSGASFPQO-----QSQ 752
DB 637	SLKLHLKKYNNHRRVPD3---IEMSGGDSPPFTSDATTFSPIMKQEIETETVEPVTSIA 693
OY 753	QSPS-----EEL-----EEL----- 760
DB 694	QPFPPPMGNMGLNANMMLKHLNDVGLMGLRNSVMSPLKCSACDPVASSADEKMRHS 753
OY 761	-----SEEEDEDEEEDV--TDEDSLAV-----RGS 786
DB 754	MSHINSSNVPTSIASLYNSLNLPSFSHVAPRDNMALESMDQVATIDDNITESHICYEEM 813
OY 787	ESGGEKAIIVRGDSEASCAEE-----EUGYVAAAATAGKEMDSN---EKTQOQSSL 835
DB 814	DQGSDSAVSPGSSQISSGDEETKCKSLSEQISAAVANGNSNPSMSANDSAMEDDESAD 873
OY 836	PPPPPPSLDOPQPMEOGSSGYLGKKEGKPRRSPSPAS--ALTPPE--GEATSVTLVEEL 892
DB 874	APHSSDITTSVSP-----PLHSSSIIVAIPIPTPONEFLQSLIAQAS 916
OY 893	SLQEMARKPEGESSSRKACVCGQAFPSQALAEHOKTHPEKPLEFTVFCROGFLEAT 952
DB 917	LLGPLLANRP-----SAFYCDHCKIPFDTLQOVLDSHMFMHTGNGF--FMCSDOQYQAFNELS 971
OY 953	LKKHMLLAHHQ 963

Db 972 FALHMYQARHQ 982 |||
RESULT 21
G01763
atrophin-1 - human
C:Species: Homo sapiens (man)
C:Date: 21-Dec-1996 #sequence_revision 06-Jun-1997 #text_change 17-Jul-1998
C:Accession: G01763
R:Margolis, R.L.
Submitted to the EMBL Data Library, March 1995
A:Reference number: G08343
A:Accession: G01763
A:Status: Preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1184 <MAR>
A:Cross-references: EMBL:U23851; NID:9915325; PID:9915326
C:Genetics:
A:Gene: GDB:DRPLA: B37
A:Cross-references: GDB:270336; OMIM:125370
A:Map position: 12p-12p
Query Match 5.5%; Score 291.5; DB 2: Length 1184;
Best Local Similarity 21.7%; Pred. No. 1.1e-06;
Matches 233; Conservative 93; Mismatches 361; Indels 387; Gaps 51;
OY 55 STDPVWVIIGQENPNNSASSEPRP-----GHNNPOVMDTEHSNPPDSSSVP----- 105
|||
Db 167 SPQPP-----DSTRQPEASFERNPSVTPTGYHAPMEPTSRMFOAPRGAPRPHPOLY 219
|||
OY 106 -----TDPTWGPERRGEESGHFLVAATGTAAGGGG-----LILASPKLGA 147
|||
Db 220 PGGTGCVLSGPRMGRKGGGAAS-----VGGPNGCKQHPPPTPTTSVSSSGASGA 269
|||
OY 148 TPLPPESTP-----APP-----PPP-----PPPGVGS-----GHLN 176
|||
Db 270 PPTKPPPTTPVGGGLPSAPRRANRPHVTNLPRLRALPLNNASASPRGAGAPRLGHLR 329
|||
OY 177 IPLLLELRVLQQRQIHQOMQTEQICROYLLLSIGQTVG-----APASSEL 224
|||
Db 330 SP-----HAM-----GQIGIGLPRGPEKPTLPAP-SPHSL 358
|||
OY 225 PGTGTASTKPLPLPFSPIKIVQTSKTLASSSSSSSSSGAETPKQ---AFPLHYPLG 280
|||
Db 359 P--PASSSAPAPRPMRFYSSSSSSSAASSSSSSSSSSASPASQALPSYRHSFP-- 413
|||
OY 281 SQPPTFASGVGRSHKPTPAPSPALPGSTDOLIASPHLAFPTTGLLAOCIGAARLEAT 340
|||
Db 414 --PTS---LSVSNQPRKYTQPSLP---SQAVWSGGRPPPPRYGRLLA-----NSN 455
|||
OY 341 ASAGGLKPKNGSGELSGEIVNGPLEKRGGRHKCRFCAKVFSGSALQIHLSHTGER--- 397
|||
Db 456 AHNGRPPRSYCGAOSTAH---PRVSTNNHHNQO-----QOQOQOQOQOHHNGSGPR 504
|||
OY 398 -----PYKCNVCGRFTTTRGNLKVHFNRHREKURYHQMNPAP-REHL-----DIVITSSG 447
|||
Db 505 PRGAFPH-----PLEGSSHHNHPYAMSPSLGSLRKYRPGRALPLRPHNSQVSYQA 555
|||
OY 448 LPYGMSVPRPEKAEBAAT-----PGGEYERRPLVASTALATES--- 487
|||
Db 556 GPNQ---PRVSSSSNSSSSSTSGSYRCSHPSPSQGQAGAPYPPRPVTVTTSSATLSVI 612
|||
OY 488 LTLTSTANG--TATAPGLPAPRK-----FVLKAKVPEKNADE-----NTPPSEGS-- 532
|||
Db 613 ATVAASSAGYKTASBPGRPYGKRAPSPGAYKTATPPGYKPGSPSFRRTGTPGRGTSR 672
|||
OY 533 -ATSGVASSATATLMQSLKMLTSLPSWALLTNHFKSTGSPPLPLCARALGASPESTKIQ 591
|||
Db 673 PAGGCTKKPGSPYGP-GLPRLPAGS-----GLPDLRPPPAASGPRLSATQIKQ 722
|||
OY 592 QLVKKIDRGAAVAVTSAASGAPTTNAPADSSSASGPNOCVTCIRVITSCPRLALRHYGH 651
|||

Db 723 EPAEYE-----TPESPVPAPRSPSPPP-----KVVDVP-----SH 753
OY 652 GGERPKCKYCGAFSTRGNLRAHFVGHKASPAARAQNSCPICQKKFTNAVTLOQHYRM 711
|||
Db 754 ASOSARFNKHLDRGFNSCARSPLYFVPLEGSKLAKR-----ADLVKRYRE 800
|||
OY 712 LGQIIPNGGTALPEGGAOENGSEQSTVSGAGSPFOQSQPSPPEELSEEEDEDEE 771
|||
Db 801 -----AEQAREREKEREREKEREREKER 825
|||
OY 772 EEDVTDSDSLAGRS---ESGGEKAISVRGDEESASGAEEYGVTAATAGKEMDSNE- 827
|||
Db 826 ERELERSVKLAQEGRAVECPSLGVRPHRPPPEGS-----AVATVPYLLGPDTP 875
|||
OY 828 --KTQOSSLPPLPPPPSLDPP--QPMEGSSGVLGKREGEKPERSSPSPALTPRGEA 883
|||
Db 876 ALRKLSTYARPHVWSPGKRNPFFVPLCAVDPLGLG---YVNPALYSSPPAARERREA 931
|||
OY 884 TSVTLVEELSLQEAMRKREPGSSSRKACEVCGQAPPSQALAEHQTHPKRGLFTCVFC 943
|||
Db 932 -----REKDLRD--RLKPG-----FEVPRS-----ELEPLHGVGP----- 960
|||
OY 944 RQGFLEERATLKKHMLAHQVOPAPRHPQNTAALSLVPGCSSTITGSLSPR 997
|||
Db 961 -----GLDPPPRHG-----GLALQPG-----PPGLHPPR 984
|||
RESULT 22
S50832
atrophin-1 - human
C:Species: Homo sapiens (man)
C:Date: 14-Jul-1995 #sequence_revision 21-Jul-1995 #text_change 07-May-1999
C:Accession: S50832
R:Nagafuchi, S.; Yanagisawa, H.; Ohsaki, E.; Shiteyama, T.; Tadokoro, K.; Inoue, T.;
Nature Genet. 8, 177-181, 1994
A:Title: Structure and expression of the gene responsible for the triplet repeat diso
A:Reference number: S50832; MUID:95144175; PMID:7842016
A:Accession: S50832
A:Status: Preliminary
A:Molecule type: mRNA
A:Residues: 1-1104 <NAG>
A:Cross-references: EMBL:D31840
C:Genetics:
A:Gene: GDB:DRPLA: B37
A:Cross-references: GDB:270336; OMIM:125370
A:Map position: 12p13.31-12p13.3112p-12p
Query Match 5.4%; Score 286; DB 2: Length 1184;
Best Local Similarity 21.9%; Pred. No. 2e-06;
Matches 235; Conservative 88; Mismatches 363; Indels 388; Gaps 52;
OY 55 STDPVWVIIGQENPNNSASSEPRP-----GHNNPOVMDTEHSNPPDSSSVP----- 105
|||
Db 168 SPQPP-----DSTRQPEASFERNPSVTPTGYHAPMEPTSRMFOAPRGAPRPHPOLY 220
|||
OY 106 -----TDPTWGPERRGEESGHFLVAATGTAAGGGG-----LILASPKLGA 147
|||
Db 221 PGGTGCVLSGPRMGRKGGGAAS-----VGGPNGCKQHPPPTPTTSVSSSGASGA 270
|||
OY 148 TPLPPESTP-----APP-----PPP-----PPPGVGS-----GHLN 176
|||
Db 271 PPTKPPPTTPVGGGLPSAPRRANRPHVTNLPRLRALPLNNASASPRGAGAPRLGHLR 330
|||
OY 177 IPLLLELRVLQQRQIHQOMQTEQICROYLLLSIGQTVG-----APASSEL 224
|||
Db 331 SPY-----AMGQMGGLPRGPEKPTLPAP-SPHSL 359
|||
OY 225 PGTGTASTKPLPLPFSPIKIVQTSKTLASSSSSSSSSGAETPKQ---AFPLHYPLG 280
|||
Db 360 P--PASSSAPAPRPMRFYSSSSSSSAASSSSSSSSASPASQALPSYRHSFP-- 414
|||
OY 281 SQHPPSAGVGRSHKPTPAPSPALPGSTDOLIASPHLAFPTTGLLAOCIGAARLEAT 340
|||


```
OY 3 HESERSRLGVPAEPAELGDASEEDHPOVCAKCAQFTD---PTEFLAHONA-----C 54
D 49 HSNPGSS--GIDAHNPQPG--AAADAFLYKCTQCHKRPPEYQSLSEHILASEPHDKLNC 104
OY 55 STDPPVYVITIGGQENPNNSASSSEPRREGHNPNVOYMDTEHNPDPGSSVPTDPTWGP 114
D 105 GAAQPRSDADEDEGSSNMSSSRVAKSPPLASNNNSSTANANNSTSSQSM-----NN 155
OY 115 REEBSGCHFLVAATGTAGGGGGLILASPKLGATLPPESTAPPPPPPPPPVCGSGH 174
D 156 NSELAKNH-----NSANKMSPMCSPGSLTPGDLPEAQLQHPQLPOLP-----H 197
OY 175 LNIPLI-----LEELRVLOQROIHOWOMTEICROVLLIGS 210
D 198 LHAQFMAAASLAMSQARTASSPSCQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQO 254
OY 211 LGQTVGAPASPELPGTGT--SSTKRLPLFSPITKVOTSKTLASSSSSSSSSGAETP 268
D 255 -----PQOLPGSNSSVGSNSAVYDLDSAP-----RSTSSPGSTTGDLSCA--- 294
OY 269 KQAFPHLYHPFGSOHPPSAGVGSHKPTPAP--SPALPGSTDLILASPHLAFSTTGLLA 327
D 295 -----YPCMOCTASFAISREQLDQHEQLHSPCGPAAVSNVQTCRICHAFAFNVYRL-- 345
OY 328 AOCIGAARGLEATASPGILKPKNGSGELSYGEVWGPLEKPGRHKRCFAKVFSGDSALO 387
D 346 -----QRHITSHDESALLR-----KFKCECKRAKFKHHLK 377
OY 388 IHLRSHTEGEPKYCNVCGNRPFTTRGNLKVHFHHRREKYPHVOMNPHRPHLDVYITSSG 447
D 378 EHVRIHSGEKPFCDGCKRSHSGFSH-----MTSKKC 413
OY 448 LPYGMSVPEPKA-----EEEAATPGGVEYERKPLVASTALSTATESTLLSTAGATAAPG 502
D 414 ISMGLKLNRRALLKRLKSPGASASSASRRSPSDHGKGLPEQPL-----PG 461
OY 503 LP-AFNKFLVLMKAVEPKKADENTPPGSEGSALISGVAESSTATLMQSLKMTSLP,SMALL 561
D 462 LPHPMSTFASDAQVOGGSAAAPAPFP-----FHPNMNALL 498
OY 562 TNHFKSTGSPPLICARALGASP-----SETSKIQOLVEK 596
D 499 -----APPHNFMMAAAGLDPRVHPYSTORLLQLSAAGQOOREEEREEOQOQOHEE 550
OY 597 IDROGAVAVYSAAGAPITTSAPAPSSASGPMQVILCLVLSCPRALRLHYGHGGERP 656
D 551 ETDPDEPLVMDIEBPTEKEMATPPEATEATP-----IKREESREASDP 595
OY 657 FKCKVCGRAEFTSTR-----GNLRAHFVGHKASPAARAONSCPIQKKFTNAVTLQOAHVM 710
D 596 ESYRSSQOAIKQOQEPFLVAVERQTPVEEHAHYEHSHADLRCSKSGQFNHPTLYOHEKV 655
OY 711 HLGQIDPNGTALPEBGGAOENGSEQSTVSGAGSFPQOQSOQPSPEBELSEEEDEDEE 770
D 656 -LCGLIKE-----ELEQHFQOQATSFALASASEDEED 688
OY 771 EEDDYTEDSLAGRSSESGEKAISYRGDSE-----ASGAEEVGTVA-- 816
D 689 EEMDYEEBP-----ROESGERKVRVRTAINEQOQOLKQOHLSTLNAKPSDEEFMTAKRL 742
OY 817 -----ATAGKEMDSNEKTQOOSLPPPPPPDS---LDOPQPMEOGSSGVL 858
D 743 QLODRVYVQVWFQNNNSRERKMSFO--NQAAAGAAPMPIDISQASLTRDQPLD-----L 795
OY 859 GGRKEGGKPESSSSASALTPGCA-----TSVTL----- 888
D 796 SVKRDPLTPKSESSPYIAPSGEALNPEAINLSRKSTASMSPASISPSAALYFGA 855
OY 889 -----VEELSL-QEAMRK----- 900
D 856 APPSPSPNSQDSTSTRSGOAFQGLPYPLMLPMLPEALFKMRPGGDFASNNALMNSIKLP 915
OY 901 -----EPGESSSRK-----AC 911
```

```
D 916 DYRGTSLSPGSEKSRMWDSDSRISHDEFGAGVILMPKPRBRGKVEITHGNAGDPDLRYVC 975
OY 912 EYVGOAFPSQAALEERQKTHKEGRLFTCVFPCROGLKEAATLKKHMLAHNOVQPF 967
D 976 DQCKAFKAKOSILARHKYHSGQRP-YOCIECPKAFKRNHILEHKL-HSEKPF 1029

RESULT 25
T31421
C-terminal domain-binding protein ral - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 29-Oct-1999 #sequence,revision 29-Oct-1999 #text,change 07-Dec-1999
C:Accession: T31421
R:Kureyev, A.; Patuturajan, M.; Litingtung, Y.; Joshi, R.V.; Gentile, C.; Gebara, M.; C
Proc. Natl. Acad. Sci. U.S.A. 93, 6975-6980, 1996
A:Title: The C-terminal domain of the largest subunit of RNA polymerase II interacts
A:Reference number: 221024; MUID:96293459; PMID:8692929
A:Accession: T31421
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 11173 <TUR>
A:Cross-references: EMBL:U49056; NID:g1438531; PID:g1438532; PIDN:AA052657.1
A:Experimental source: h1pocampus

Query Match 5.3%; Score 279.5; DB 2; Length 1173;
Best Local Similarity 20.0%; Pred No. 4,1e-06;
Matches 242; Conservative 110; Mismatches 383; Indels 477; Gaps 46;

OY 106 TDPPTGERRRGE--SGHFVLAATGTAAGGGGLILASPKLA---TPLPSTPAPRP 160
D 72 TGTIVTQSHASRPACSRHLTLTGT---GDGAPAPAPPSGSSSPSSSSSPSPPP 127
OY 161 PPPPPPPGV-----GSGHNIPLIL 181
D 128 PPPPPPALPAPRPDIYDPFHPTEAIVSPPAPEQYDPPEATGSPSSSGTSPSEEE 187
OY 182 EELRVLOQROIHOWOMTEICROVLLIGSL-----GQTVGAPA 219
D 188 EE-----EEEEEEBGLSQIRIRHSETLAGIYDDNSLSQDFPGDDSPHNEPPPTLGAPG 243
OY 220 SPSELPGT-----GTAASKPLLP 238
D 244 TTPQADSTRABGAPRRRVVGVPEAEALCEKGVSEVYTTAGGAPALPLPLPPTDPEIEE 303
OY 239 -----LESPIKPVQTSKTLASSSSSSSSSGAETPKQAFHLYHPLG--- 280
D 304 GEIYQPEEPRVAVSLFRAARPRQPPASVATLASVAAPAPAPASAPRA-----PEGDF 357
OY 281 -SOHPFSAG---GVGRSHKPTPAPSPALPGSTDLIASPHLAFPTTGLLAOCIGANR 335
D 358 LSLHADSQEGALQYDLGPRAPAPADARKGGLD-----LRKILTORR 401
OY 336 G-LEATASPG-----ILKPKNGSELSTGYEMGLEKP---GGRHKRCFCAVFGSD 383
D 402 ERYRQRSASPGPPPARKKARRERORSQD-----PAPDSTTWEAKRRSR--ERKLGS 453
OY 384 SALQIHLRSHTEGEPKYCNVCGNRFTTRGNLKVHFHHRREKYPHVOMNPHRPHLDVYI 443
D 454 STARRSRKSRKRRS--KRSADR--RGS---HRSRSREKRRRRRRKASPP-----AA 501
OY 444 TSSG-----LPYGSVPPKAEAEATPGC 468
D 502 SSSSSRRRHRHGKRRREGGKKKKRKRRAEKRGDLEKLP--APVPPSGSDRDSRRRGA 559
OY 469 GYERKPLVASTALSTESTITLSTAGTATAPCLPAFNKVLKMAVPEPKN----- 520
D 560 ---VPPSIDLTDHDLFAIKRTITVGRDPDKTEPPAPS-----PAPAVSPKGGLVLYDSGL 611
OY 521 -ADENTPPGSEGAISGVASSTATLMQSLKMT-----SLPSWALLTNHFKSTGS 570
D 612 SADERGAKDKDRRRRGAASSSSSSSRKASRRKALDDGDRGRDRDSSSKKPRTPKQSAPOS 671
```


[illegible]

Db 76 GVGPAALITTRWRSPRGSRG--RPSTGGGVVR-----GGR--CDVCGKVYSQRS 120

OY 385 ALOIHLRSHAGSERPYKCNVCNGNREPTTGNLKVHFHRHREKYPHAVOMNPAPVEHLDYIT 444
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 121 NILRHQKIHTGERPFCVSECSGRSFRSSHLRLHQTLTHEER-----FVCG 166

OY 445 SSGLPFGMSVPPEKAEEAATPGGCVETKPLVASTALLATESLLLSLTASGTATAPGLP 504
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 167 DCGQGF--VRSARLEHRRVHTG--EQPRCAECGGSFQRSNLQHQRTHGDPG-- 218

OY 505 AFNKFPVLMAVEPKRNKDENTPGESEGSAISGVASESTATIMQLSKMTSLPSMALLTNH 564
|||::|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:
Db 219 -----PGAKPPAPPAPGPF----- 231

OY 565 FKSTGSEPIPLCARALGASPETSKTLOOLVERIDROGAVALVTSAAGAPTTPAPSSSA 624
|||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:
Db 232 --PPGFPPCECR-----ESPARAVLLEHQAVH-----TGDK 262

OY 625 SSGPNOCVICLRVLSCPRALRLHYGHGGEPPFKCKVCGRAFSTGNLPAHYGHKAS-P 683
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 263 STFG---CVEEGEERFGRRSVYLQHRRVHSGERPFACAEQGQSFRORSNLTQHRIHTGERP 319

OY 684 AAANAQSCPTCCOKFTNAVVLQOHVRMLIGQIPRGALTALPGGGAAGQNSGQSTVSGA 743
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 320 FA-----CAECGAFRORPPLITLOHRLVHTGK----PRACPFG----- 354

OY 744 GSFPQQSQDQPSEELSEEDEEDEEDVDTDSDLAGSGSESGEKAIIVRGDSEEA 803
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 355 -----QRFSGRLKLTRHQRTHT-----GEK----- 374

OY 804 SGAAEEVGIVAAAATAKGEMDSNEKTTQSSLPPPPPDLSDLDQPMEDGSSGVLGKEE 863
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 375 -----PYHCGEGL----- 383

OY 864 GGPKESSSPASALTPREGATSVTLVEELSLQEAMRKPESSSRKACVCVOAAPSOAA 923
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 384 -----GFTQVSRLT--EHORITHTGERPF--ACEPCGOSFROHAN 418

OY 924 LBEHQTKHPKEGPLTCVFCRQGLFLERATLKMHMLAHNOVOFA 968
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 419 LNQHRIHTGERP-YACRPEGKAFRQRPRTLQH-LRTRREREFFA 461

RESULT 27

Arylphorin gene-specific binding protein 2 - flesh fly (*Sarcophaga peregrina*)
JC5146
N:Alternate names: ABP-2
C:Species: *Sarcophaga peregrina*
C:Date: 31-Jan-1997 #sequence_revision 27-Feb-1997 #text_change 01-Dec-2000
C:Accession: JCS146; PC4244
R:Adachi, N.; Kubo, T.; Natouri, S.
J: Biochem. 120, 1239-1246, 1996
A>Title: Purification, characterization, and cDNA cloning of ABP-2 (*arylphorin* gene's aga peregrina).
A:Reference number: JCS146; MUID:97164029; PMID:9010776
A:Accession: JC5146
A>Status: Preliminary
A:Molecule type: mRNA
A:Residues: 1-594 <ADA>
A:Cross-references: DBDJ:D86231; NID:g1753198; PIDN:BA13048.1; PID:g1753199
A:Accession: PC4244
A>Status: preliminary
A:Molecule type: protein
A:Residues: 150-157/193-204 <AD2>
C:Comment: This protein is a negative transcription factor of the arylphorin gene.
C:Superfamily: zinc finger protein ZFP-36; LIM metal-binding repeat homology
F:246-251/Region: alanine-rich
F:258-271/Region: glutamine-rich
F:401-411/Region: glycine-rich
F:413-420/Region: histidine-rich
F:480-496/Region: asparagine-rich

Query Match 5.3%; Score 277.5; DB 2; Length 594.


```

021 821 -----KEMDSNEKT-TOQSSLPFPPPPPSDLPQPPQMEGSSGVLGKEGSGKP 867
022 951 DHIRHFTDSLIMKREMEENOYTDALSSISSHVPPELKIOTLHRK-----SKSOAYAM 1002
023 868 ERSSSPASALTPEGGATSVTLVEELSLQEAARKKEGESSS 907
024 1003 MSLSDKDSLHPTSHSSS-----NYWHSMARAAAESSA 1035
025
026 RESULT 31
027 zinc finger protein ZNF43 - human
028 S26823
029 N.Alternate names: zinc finger protein kox27
030 C.Species: Homo sapiens (man)
031 C.Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 01-Dec-2000
032 C.Accession: S26823; 137967; S10416
033 R.Loverling, R.; Trowsdale, J.
034 Nucleic Acids Res. 19, 2921-2928, 1991
035 A.Title: A gene encoding znc2 highly related zinc fingers is expressed in lymphoid cell
036 A.Reference number: S26823; MUID:91279444; PMID:1711675
037 A.Accession: S26823
038 A>Status: Preliminary
039 A.Molecule type: mRNA
040 A.Residues: 1-803 <LOW>
041 A.Cross-references: EMBL:X59244; NID:g38031; PIDN:CAA41932.1; PID:g38032
042 R.Thiesen, H.J
043 New Biol. 2, 363-374, 1990
044 A.Title: Multiple genes encoding zinc finger domains are expressed in human T cells.
045 A.Reference number: 137949; MUID:91145339; PMID:2288909
046 A.Accession: 137967
047 A>Status: Preliminary
048 A.Molecule type: mRNA
049 A.Residues: 476-531 <THI>
050 A.Cross-references: EMBL:X52358; NID:g34160; PIDN:CAA36584.1; PID:g930090
051 A.Gene: GDB:ZNF43; HTF6
052 C:Genetics:
053 A:Cross-references: GDB:128653
054 A:Map position: 19p13.1-19p12
055 C:Superfamily: zinc finger protein ZFP-36; LIM metal-binding repeat homology
056 C:Keywords: DNA binding; zinc finger
057
058 Query Match 5.2%; Score 275; DB 2; Length 803;
059 Best Local Similarity 20.1%; Pred. No. 4.5e-06;
060 Matches 135; Conservative 59; Mismatches 194; Indels 282; Gaps 22;
061
062 347 KPNKNGSELSTGYEWMGLEKPKGRKRCFKACAVFGSDSLQIHLNRSHTGERPKVCNVCN 406
063 287 KANQSSNLTEHKKIHNGEKP--YKCECGKAFWMPSTLTIKRIRHTGKPKYTCCEGK 343
064 407 RFTTRGRLKLVHNRH-REKYHNVQMNPRVREHLDYVLTSSGPRYGMNVPRKEAEDEAT 465
065 344 AEFQFSULTTHKRIHTEKFP-----YKTECG----- 370
066 466 PGGGERKPLVASTATASATESLTLLSTAGTATAGLPFAKFKVLMAKAVEPRKKADENT 525
067 371 -----EARSRSNSLT-----KHKKIHEKK 390
068 526 PGGSE--GSAISGVAESSTALIMOLSKLMTSLPSW-ALLTNHFKSTGSPFLPICARALGA 582
069 391 PYKCECGKAF-----KWSKLTEN----- 410
070 583 SPSETSKLOOLVEKIDQGAVALVTSASGAPTTSAAPAPSSSASSGNOCVIGLRVLSCLR 642
071 411 -----KLTHTGER-----PYKCECGKAFWMP 433
072 643 ALRLHYGOHGGERPFCKCKVCGRAFSTRGNLRAHFWGHKASPARAONSCDICOKKFTNAV 702
073 434 TLTKHNRIHNGEKPCKEYCGKAFQFSNLT-----HKRIHTEKPKYKCECGKAFSRSS 489
074 703 TLDOHVMHMLGQGLPNCGTALRPEGGAAGQENSGDSTVSGAGSFPOQOSQSPSEELSE 762
075 490 NLTKKKIKIH-----EKKPYKCECGK 511

```

```

OY 763 EEEEDDEEDDYDDEBSLAGRSSE-----GEAIVRGDSEFASG 805
Db 512 AFKWSKLTHTKTIHTGEKPYKCECGKAFFNHFSILTKHRIHTGEKPYK----- 562
OY 806 AEEEVG---TVAAATATGKENDSNEK-----TTOOSSLP PPPDDLDOPQMEQ 852
Db 563 --EECGKAFQSSNLTHHKIHTGEKPYKCECGKAFTQSSNLTHHKIHT----- 611
OY 853 GSSGVLGKKEEGKPERSSSPASALTPEEATSVTLVEELSLQAMKKEPESSRKACE 912
Db 612 -----GKPYKCECGKAF---NOSTLTTHKIIHTGEKPYK-----CE 647
OY 913 VCGQAFPSQAALDHQKTHREGEPLTCYFCQNGFLERATLKKHMLLAHQVPP----- 967
Db 648 ECGKAFKWSSTLTTHKIIHTGEKPYKCECGKAFTLSLTTHKIIHTGEKPYKCEKC 705
OY 968 --APHGPPQNI 975
Db 706 GKAFNRPNSLT 715

RESULT 32
T13049
eyelid - fruit fly (Drosophila melanogaster)
C:Species: Drosophila melanogaster
C:Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 17-Nov-2000
C:Accession: T13049
R:Rietsman, J.E.; Luk, A.; Rubin, G.M.; Heberlein, U.
submitted to the EMBL Data Library, March 1998
A:Reference number: 217592
A:Accession: T13049
A>Status: preliminary; translated from GB/EMBL/DDb
A:Molecule type: mRNA
A:Residues: 1-2715 <TRES>
A:Cross-references: EMBL:AF053091; NID:g2981220; PID:g2981221; PIDN:AAC06254.1
C:Genetics:
A:Gene: eld
A:Cross-references: Flybase:FBgn0003013
C:Function:
A:Description: could act as a transcription factor antagonistic to the Wg pathway
C:Keywords: DNA binding

Query Match 5.2%; Score 275; DB 2; Length 2715;
Best Local Similarity 21.7%; Pred. No. 1,7e-05;
Matches 220; Conservative 88; Mismatches 345; Indels 360; Gaps 51;

OY 13 VPAGEPAELGDAEEDHPRQVCAKCCAQFTDPTETLANHQAACSTDPPVMVITGGEN--P 70
Db 166 LPGKKRPQ-----QQQPHQ-----QQPQQPGGGSPNRRP 197
OY 71 NNSASASPREGHNHPQVMDTEHSNPRDSSVPTDPTWGPERGESSGHFLVAATGT 130
Db 198 QQRYPGPQ--PQRTPTLNSLQSSNPPRRPQRHYAN--TYDQQA--ASAAAAAAQOO 253
OY 131 AAGGGGGLTASPRLGATPLPESTPAP-----PPPP-----PP-- 166
Db 254 QAGG-----PPRGHGRPRQNHQPSRYCGQGGKMARPPRRYRVSQQLRPPRPIN 306
OY 167 -----PRGVSGHNLPLILELRYLQORQIHQMOTQIGCRQVLLGSLGTGAP 218
Db 307 TSKGQSYPRPANHONGSGSYR-----SSPQQOQOQOQOQOQOQO-----AGQDPGGP 350
OY 219 ASRSELPGTG---TASSMKPL-----LPLFSPIKR-----VQTSKTLASS----- 255
Db 351 VPGGRPRGTGQORPQQNRPRTSQSYPRQRYPRRGLRAGGSNHTAYSTHOYRPPNRPM 410
OY 256 --SSSSSSSSGAETPRKQAFNLHNPILGSONP-----FSAG-----VGRSHKTPAPSPA 303
Db 411 PGGSFPRGSGHPLRPASNNH--RPLQOQRRRRPNVAGSGRRPSSSPGNARPSQQA 469
OY 304 LPGSTDILASPHLAFRSTTGLLAADQLGARGLEATASRGLKPKNGSGELSTGEYMG 363

```


Db 470 SSPHOELIGON--SNSSSSG-----GAHSGM-GSGPGRPNDD-----QVMRP 510

QY 364 LEKPGGRHKRCFAKAYVGSBALSALDILHLSHTGERRYKCAVCNGNRFTRGNLAKYHFRHRE 423

Db 511 TPSPPTGSS-----GSRMSMSPVAONHPISRP-----ASNSSSGGPM-----Q 548

QY 424 KYRHYOMNHPVPEHLDVYITSSGLPYGMSVPPPEKAEFEAAAPGGGVERRKPLVASTALS 483.

Db 549 QRPVGAAGGPRPMPRH-----PGMPG---PPQOQOQSQOQO-----ASNASS 587

QY 484 ATESLTLLSTACTATAPRGLPAFNKFLVLMKAVEPKKADEN--TPPGEBSGA-----533

Db 588 ASNS-----PQOQPRPPAP-----PPNOGNNNNATPPPPQGAAGGGMMP 627

QY 534 -----ISGVNESSATATLMQSLKMTSLPMSMALLTNFKSTGSP---LPLCARALG 581

Db 628 PHMHGQKMGPGQSPQAQGYPPQOPOOYP-----GNTPPRQYPRGAYATG 675

QY 582 ASPSETSKLQOLVERKIDROGAVAVTSAASGAPTTAPR--PSSASASPNOCVICLRVL 638

Db 676 PPPPPTS-----QAGAGANSMPSGAQAGGYPGRGMPNHTGQYPPYQVPPSPQ 725

QY 639 SCPRALRLHYGQNGGERPFCKYCGRAFTSTRGNLRAHYVGHKASP---AARQNSCPIC 694

Db 726 TVPG-----GADGG-----AMVGNHYQSGKTPRPVYGGPPPGSGSPRLN-----768

QY 695 OKKFTNAVTLLOQHVRMLHLSGOIPNGGTAALPEGGGAAGSSEOSTYVSGAFSPQOOSQP 754

Db 769 -----YIKQHQL-HKGGY---GGSPTPPQPGCYGNGP-----TGMHGMMPMP 809

QY 755 -----SPEELSEEEDEDEEEDVTDDEDSLARGSESGEKAISV 796

Db 810 HHMGPRHGTNMGPRPTSTPPQOSQILGGQPPQG-----ASGGPESGPREHIS- 857

QY 797 RQDSEASGAEEVEVG--TVAAATAGKEMDNSEKTKOOSLPP-----838

Db 858 QDNGISSSGPTGAAGMAHYVSVYTTGPDGTSMDEVSOQSTLSNAAASGEDPQCTTPKSR 917

QY 839 -----PPPDSLDQPPMEQSSGVLGKKEGKGPENSSSPASALTPEG 881

Db 918 KNDPYQSHLAPSTSPHPVPMHPG-----GGP---GGEYDMSSPMPMRPAG 962

RESULT 33

A49073

HSN motor neurons migration regulator (alternatively spliced) Egl-43 - *Caenorhabditis elegans*

C:Species: *Caenorhabditis elegans*

C:Date: 07-Apr-1994 #sequence_revision 18-Nov-1994 #text_change 01-Dec-2000

C:Accession: A49073; B49073; T24250; T24249

R:Garriga, G.; Guenther, C.; Horvitz, H.R.

Genes Dev. 7, 2097-2109, 1993

A:Title: Migrations of the *Caenorhabditis elegans* HSNs are regulated by egl-43, a gene e

A:Reference number: A49073; WUID:94040732; PMID:8224840

A:Accession: A49073

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-581 <GAR>

A:Cross-references: GB:S66757; NID:g452950; PID:g452951

A>Note: sequence extracted from NCBI backbone (NCBI:139438, NCBIP:139439)

A:Accession: B49073

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 222-581 <GA2>

A:Cross-references: GB:S66936; NID:g452952; PID:g452953

A>Note: sequence extracted from NCBI backbone (NCBI:139899, NCBIP:139727)

R:Wilkison, J.

submitted to the EMBL Data Library, October 1995

:Reference number: Z19863

:Accession: T24250

:Status: preliminary; translated from GB/EMBL/DBJ

:Molecule type: DNA

:Residues: 1-537, 'VPVP', 542, 'V' <WIL>

:Cross-references: EMBL:566515; PIDN:CAA91353.1; GSPDB:GN00020; CESP:R53.3a

[illegible]

Db 338 PEDFLKASLAYEMERTYI-----THSPLPSSTT--PSPPASSPEQSLKSSSPHSSP--- 388
Oy 101 GSSVPTDPTWGBERRESESGHFLVATGTAAAGGGGGLILASPKGATPLPESPAPPP 160
Db 389 GNTVSPRLAPGLPEHRDSYS---YLVNSYG-----SEGIGSY---PGYAPAPHL 430
Oy 161 PPP-----PPPGVSGHINIPILILELRLVLOQRQIHQMOMTEOICRQVLL 208
Db 431 PPAFTPSYNAHNPFLPLPGISSNGLS---TMNNININNSL--FRLPYPVSNLLSG 485
Oy 209 GSLGDTGAPAS-PSELPGTGTASTKPLPLPFSPIKPVQTSKTLASSSSSSSSSGAET 267
Db 486 SSLPRLMPLASPLSSLPTDG---ARRLLP-----PEH 515
Oy 268 PQQAFPHLYHPLGSSQHPFSAGVGRSHKPPAPSPALPGSTQDLASPLAP-STITGL 326
Db 516 PREVLIRAPH-----SAFSLTGAASMKDESSPPSGSP-TAGTAATSEHVQKATSSVM 569
Oy 327 AAQCLGAAGLEATASPLGLPKNGSELSTYGEVWGPLEKPGG--RHKCFCAKVFSGDS 384
Db 570 AARPTDGAANL-----INKRMWGTGYKTLRPLKKGOKIKYEONCAKFTGOLS 619
Oy 385 ALQIHILRSHTEGRRPYCNVCGNFRFTTGKLVNHEHRHREKYPHVQMPHPVPEHLDTYT 444
Db 620 NLKVHLRVHSGERPRFCQTCNKFTQLANLQKHVLYHTGEKPH----- 662
Oy 445 SSGLPYGMNVPRPKAEBAATPGGVERKPLVASTTALSATSLTLLSNAGTATAPGLP 504
Db 663 ----- 662
Oy 505 AFNKFLMKAVERPKNADENTPPGSEGSALSGVAESSTATLMLSKLMTSLPMSALLTNH 564
Db 663 ----- 662
Oy 565 FKSTGSPRLPLCARALGASPSETSKLOOLVEKIDROGAVAVTSAASGAPTTAPABSSA 624
Db 663 ----- 662
Oy 625 SSGPNCVCILRYLSCPRALRHYGONGGERPKKVCGRATSTRGNLAHFHGHKASPA 684
Db 663 -----RCQVCHKRRPSTSNLKTHLRLHSGEKPYOCKVCRAKTFQVHLKL---HKRLHT 713
Oy 685 ARAQNSCPICQKFTNAVTLOQHVHMLGQIIPNGGTALPEGGGAQENSGSTVSGAG 744
Db 714 RERPHKACQCHKSVYIHCSL---KVHLKGNCRAG-----PAAG----- 748
Oy 745 SFPPQOSQOPSPPEELSEEBE-----EDEEEEDVTDEDSLGRGSESGE 791
Db 749 -----LPLEDLTRIMEIERFRDISDNADRLMEDSDVDT-----SMVE 787
Oy 792 KAI--SVRGSESEAS--GAEEVGT--VAAAAATAGKEMD-SNEKTTQSSSLPPPPPS 843
Db 788 KETLAVAVRKEKETSLKVSLOQRMNGNGLSSGCSLYESSDLSLMLKPHSNPLPLVPVKV 847
Oy 844 LDQOPDME 851
Db 848 QETVERPMD 855
RESULT 35
E88280
protein egl-43 [imported] - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 10-May-2001
C:Accession: E88280
R:anonymus, The C. elegans Sequencing Consortium.
A:Title: Genome sequence of the nematode C. elegans: a platform for investigating biolog
A:Reference number: A75000; MUID:99069613; PMID:9851916
A:Note: see webistes genome.wustl.edu/gsc/C_elegans/ and www.sanger.ac.uk/projects/C_elg
A:Accession: E88280
A:Status: preliminary

A:Molecule type: DNA
A:Residues: 1-543 <STO>
A:Cross-references: GB:chr.II; PIDN:CAA91353.1; PID:93879174; GSPDB:GM00020
A:Note: similar to zinc finger, C2H2 type (4 domains)
C:Genetics:
A:Gene: egl-43
A:Map position: 2
Query Match 5.0%; Score 264.5; DB 2; Length 543;
Best Local Similarity 25.7%; Pred. No. 9.4e-06;
Matches 108; Conservative 59; Mismatches 191; Indels 63; Gaps 14;
Oy 350 NGSGEL--SYGEVWGPLEKPRGRHKRCFAKVGSDALQILHRSHTGERPYKCNVGNR 407
Db 136 NGEEETIREHGLRPGQTPRGDSHKCVCKPSFSSASGLKQSHINCSLKPFGHLCPKS 195
Oy 408 FTTGNLKVHFNHREK-----PHVQMP--HPVEHLDYVTTSS 446
Db 196 YTOFSNLCRRHRRVSDWTCPTCOSQMPQALTRHRYCENTALYKPLMAOLAGLSGAG 255
Oy 447 GL-----PYGMSVPRPKAEBAATPGGVERKPLVASTTALS--ATESLTLTSTAGTA 498
Db 256 GLGSPRYPHILOMATQRYNPLFLANPREAYKLMQOTTCASPRDACSCHASSESPPT 315
Oy 499 TAP-GLPAFNKFLMKAVERPKNADENTPPGSEGSALSGVAESSTATLMLSKLMTSLPS 557
Db 316 TEEDVDLATPKRPTSEMETTSKSDGDHSDSDGNDDDSEAGVLDESSTTTS--- 372
Oy 558 WALLTNHFKSTGSPRLPLCARALGASPSETSKLOOLVEKIDROGAV---AVTSAASGAP 613
Db 373 ---TKKRPYSHITSDILAARQALNSTPLGMLORSILNYPVPSPHFLRAMSGAK 428
Oy 614 TTSAPAPSSASGPNOCVILRYLSCPRALRL--HYGONGGERPKKVCGRATSTRGN 671
Db 429 ASSPSSSSGSGKDRYTCKFCQKVF--PRSANLTRLHLRHTGTGOPYKQCCENSFSISSN 486
Oy 672 LRAHF--VGHKASPAARAQNSCPICQKFTNA-----VTLOQHVHMLGQIIPNGGTAL 723
Db 487 LQNHVRINHKNRPNSLPRHNHNR--QRLHNSTSTSTTTTVHNPRLHL-----PQTSV 538
Oy 724 P 724
Db 539 P 539
RESULT 36
S06546
finger protein (clone X1COF.1) - African clawed frog (fragment)
C:Species: Xenopus laevis (African clawed frog)
C:Date: 28-Feb-1990 #sequence_revision 28-Feb-1990 #text_change 01-Dec-2000
C:Accession: S06546
R:Nietfeld, W.; El-Baradi, T.; Mentzel, H.; Pieler, T.; Koester, M.; Poetting, A.; Kno
J. Mol. Biol. 208, 639-659, 1989
A:Title: Second-order repeats in Xenopus laevis finger proteins.
A:Reference number: S056332; MUID:90040698; PMID:2509712
A:Accession: S06546
A:Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-615 <NIE>
C:Superfamily: zinc finger protein ZFP-36; LIM metal-binding repeat homology
C:Keywords: DNA binding; zinc finger
Query Match 5.0%; Score 264.5; DB 2; Length 615;
Best Local Similarity 21.4%; Pred. No. 1.1e-05;
Matches 134; Conservative 61; Mismatches 237; Indels 195; Gaps 21;
Oy 365 EKPRGRHKRCFAKVGSDALQILHRSHTGERPYKCNVGNRFTTGKLVNHEHRREK 424
Db 140 EKP---FCCSPDCDCFASSSELNHOQRTHGKPKFSCSECGKPTNHS---HFAHNO-- 190
Oy 425 YPHVQMPHPVPEHLDYVTTSSGLPYGMSVPRPKAEBAATPGGVERKPLVASTTALSA 484
Db 191 -----MTHNG-----EKPRCCKGCKCF 208


```

y 485 TESLLTLLTSAGATATAPGLPAFNKFLVLMKAVERPNKNAKDENTPPGE--GAISGVAASSSTA 543
b 209 ASSSDL-----TFNR-----RTHREKFTFSCECGKCFNSHSHLARR 245
y 544 TLMQ-----LSKIMTSLPSMALLTNHFKSTGSPPLPLCARLALGASPESTSKL--QQLV 594
b 246 QMIHNGEKFPCSECGKCFSSSSSLTAHQRTNHKKVPRFSCGACGKCFENRSLIHTHOMT 305
y 595 EKIDROGAVAVTSAASGAPTTAPAPSSASSGPNQVICTLVLSCPRALRLHYGGHGE 654
b 306 HTGK-----PFSCFCEKRCFSPNSMLARHOMHTGE 337
y 655 RPFCKVCGRAPESTRGNLRAHFVGHKASPARAONSCIPQCKKFTNAYVLQOHVRNHLGS 714
b 338 KPFSCSECKCFSSSSDLTFNHRHT---TGKPRPSCGCGCYKSKSLVNHQRTHT-G 392
y 715 QIPNGTALPEGGAOENSGEOSTVGAGSFPOOOSOPSPHEELSEEEDEDEEED 774
b 393 EKPFSGSKDKDCASSSELNIHQRTHTGEKAF-----SCSECGKC 432
y 775 VTDEDSLARGSESGEKAISVRGDSBPASGAEBEVGTVAATAAGKEMDSNEKTTQSS 834
b 433 FTNRSOLSRHOMHTGEKPIISC--PECEECF-----VSSSLTAHQOQ 473
y 835 LPPRPPLDQQRPMHQSSGVLGKKEEGK--PERSSPSPALPBEENAVTLVEEL 892
b 474 AHRMYKPFCC-----LECGKCFSPNRRFARHOMHTGE-----KPR 509
y 893 SLQEMARKPEGSESSSRK-----ACEVCGAFPSQALAEHOKYHPRKGPFLTCV 941
b 510 SCSECKGKFSNOSSLARHOMHTHNGEKFPCSECGKRFENOSILARHOMHTGKRP-FSCS 568
y 942 FCRGQFLERATLKKHMLLAHHQVPR 968
b 569 ECAKGFSNOSGLARHQ-MTHTGKPRFA 594

RESULT 37
\Incogene: Evi-1 - human
\Species: Homo sapiens (man)
\Date: 20-Feb-1993 #sequence_revision 17-Apr-1993 #text_change 05-Nov-1999
\Accession: A60191
\Morishita, K.; Parganas, E.; Douglass, E.C.; Ihle, J.N.
\Title: Unique expression of the human Evi-1 gene in an endometrial carcinoma cell line
\Reference number: A60191; MUID:90326419; PMID:2115646
\Accession: A60191
\Molecule type: mRNA
\Residues: 1-1051 <MOR>
\Cross-references: GB:X54989; NID:g50873; PIDN:CMA8735.1; PID:g50874
\Note: the authors translated the codon CCT for residue 85 as Leu
\Genetics:
\Gene: GDB:EV11
\Cross-references: GDB:119889; OMIM:165215
\Map position: 3q26-3q26
\Keywords: alternative splicing; DNA binding; zinc finger

Query Match 5.0%; Score 264.5; DB 2; Length 1051;
Best Local Similarity 18.7%; Pred. No. 1.9e-05;
Matches 171; Conservative 84; Mismatches 236; Indels 425; Gaps 34;

y 368 GGRHKCRFAKVTGSGSALDIHLRS-HTGERPYKCNVCGNRTTGNLKVHPRH----- 421
b 128 GKHVEECNCAKVTDPDSNLRHNRHSOHVGARAHACPECGKTFATSSGLQKHKIHSSVKP 187
y 422 -----REKYPHY-----OMNPH-PVPEHLLDYVIT 444
b 188 FICEVCNKSYYTQSNLCRRKRRMHADCRITQIKCKDCGQMFSTTSLNKKRRFCBGKHHFAA 247
y 445 SSGLPYGMGVPRPEKAEDEAAT----- 465

```

Db 248 GCF6FGGIGISLIPGPRPAMDKTSVMNMSHANPGLADYFGANRHRPAGLTPPAPGFSFSPGFL 307
 QY 466 PGGVEKRPVASTT---ALSATESLLTSTAGT-----ATADGAPFNKFLVM---K 513
 Db 308 PSLYHRPPLIPASSPVKGLSTSEQTNKSQSLPMTHPOLIPATODILKSLKSPVGDNK 367
 QY 514 AVE--FNKNDE-----NPPSE-----GSAISGAESTATLMOL 548
 Db 368 PVELPDRSRSEERPEEKISDQSESSDLDVSTPSGSDLETTSGSDESIEDSEKKEFKEN 427
 QY 549 SKL-----MTSLPSSWALTIN-----HFKS 567
 Db 428 GKMFKXKVSPLQNLNAGINNKKKEYSNHSIFPSLEQGTAVSGAVNDSIKAIASIAEKYFGS 487
 QY 568 TG-----SPLP-----LCARALGASBETSKLQ 592
 Db 488 TGLVGLDQKKVGLALPYPSMPLPFPFPAPFSQSMYPPDRDLRLPLKMEQSPGEVKLOK 547
 QY 593 -----LVEKIDRCGAVAVTSAASGAPPTTSAPAP-----SSGASSG----- 627
 Db 548 GSSESEFDLTTKKRDEKPLRPVPSKPPYTPATSDQPLDLSMGRSRAGTKLTERRKNH 607
 QY 628 ----- 627
 Db 608 VFEGKKGSNYESRPSDGLQHARPTPEFMDPIYERKERRKLTPLPLEALKELYLRPSGFL 667
 QY 628 -----PQ----- 630
 Db 668 PHPQPLDPDRTPMWSAIEENNAEKLSEFSALKPEASELLOSVSMFNFRAPRNALPENLR 727
 QY 631 -----CVICLRVLCSPRLRL--HYGQHGGERPCKKVCGRAPSTGNLRAHF--VGH 679
 Db 728 KGERTKRFGCGKIF--PRSANLTRHLRTHTEGQPYRCKYCCRSFESSNLRQHVNIHN 785
 QY 680 KASPARAQNQSCPTCKKFTNAVTLQGHVRLHGGQIINGGALP---EGGAA----- 730
 Db 786 KERP-----KCHCYQCFGGQTNLDRLHKLKHNENMSGTASPSSELESTAILDK 840
 QY 731 -----QENSEE-----OSTVSGSPFOQSQSPSEBELSEEBEDE--- 769
 Db 841 DAVFTEIRNFTGNSNHSQSPRVNVEERNNGS-HFKERKALVPSQNSDLDEVEDEVELL 899
 QY 770 -EEEDV-----IDEDSLAGSGESGGEKASIVRDSERASG-----ABEEVG 811
 Db 900 DEDEDEYDITGKTGKEVNTSLNHEGNPE-----DYEFTSALNESCCKTSPVRYKKEEYK 953
 QY 812 TVAA-----ATAGKEMDSNE-KTQOSSLPPPPPSDLDPQPEQSGSV---- 857
 Db 954 SGLSALDHITHFTDLSLKMKNKMDNQYSBALSSFTSHVPEELKQPLHNRKSNQAVAMML 1013
 QY 858 -LGGKEEGKPERNSS 872
 Db 1014 SLSDKESLHSTSHSS 1029
 RESULT 38
 S03677
 finger protein (clone mkr3) - mouse (fragment)
 C:Species: Mus musculus (house mouse)
 C:Date: 04-Dec-1992 #sequence_revision 04-Dec-1992 #text_change 01-Dec-2000
 C:Accession: S03677
 R:Chowdhury, K.; Rohdewald, H.; Gruss, P.
 Nucleic Acids Res. 16, 9995-10011, 1988
 A>Title: Specific and ubiquitous expression of different Zn finger protein genes in t
 A:Reference number: S03677; MUID:89057528; PMID:3143103
 A:Accession: S03677
 A:Status: not compared with conceptual translation
 A:Molecule type: mRNA
 A:Residues: 1-428 <CHO>
 A:Cross-references: GBX:112592; NID:g53132; PIDN:CA31105.1; PID:g53133
 A>Note: The sequence from fig.2 is inconsistent with that shown in fig.1 in having 34
 C:Superfamily: zinc finger protein ZFP-36; LIM metal-binding repeat homology
 C:Keywords: DNA binding; zinc finger

Best Local Similarity 18.6%; Pred. No. 1.6e-05;
Matches 155; Conservative 90; Mismatches 248; Indels 341; Gaps 26;

```
QY 44 PNEFLHQNCSTDPFMYVILIGGQENPNSSASSPREGHNHPQ---VMDTEHSNPDS 100
| : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 270 PEDFLKASLAYGIERTPTI---TRSPISSTP--PSPARSSPDOSLKSSSPHSPGNT 323
| : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 101 GSSVPTDPPTWGPERRGESSGHFLVAATGTAAGGGGLILASPKLGATP-LPPESTPAP 159
| : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 324 VSPV-----GPSQEHDRDSAYALNASYGTGELG-----SYGVAPLPRLPPAFITSYN 371
| : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 160 PPPPP--PPPPGVSGHLNPLLEELRYLQQRQHOMQTEQIC--ROYLLIGSLGQT 214
| : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 372 AHYPKFLPPLPYGMNCGLS-----AVSSMNGINNFGLFPRLCPVYSNLLGGSLPHR 423
| : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 215 VGAPAS-PSELPGTASTPKPLPLFSPIKPVQTSKTLASSSSSSSSSGAETPKQAF 273
| : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 424 MUNEPTSLPSSLPDGAARLLQPEHP-----REVLPAPHSASFSTGA----- 465
| : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 274 HLYHPLGSOHPFAGGVGRSHKPTAPAPSPALPGSTDOLIASPHLAFPTGLLAQCIGA 333
| : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 466 -----AASMKDKACSPTSGPSPTAGTAATAEHVYQ---KATSAAMAAPSSDE 509
| : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 334 AKGLEATASGLLKPKNGSGELSGEVMPLEKPCG--RHKCRFCAKVSGDSALQIHLR 391
| : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 510 AMNL-----IKKRRMTGKTLPLPYPLKKQNGKIKYECNCAKTFGQLSNLKVHLR 559
| : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 392 SHTGERPYKCNWCGNRPFTGNLKVHFHRRHREKYPHVQNMNPPRPHLDVITSSGLPYG 451
| : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 560 VHSGERPFKQCTCNKGTQLAHLQKHLYVHTGERKH----- 595
| : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 452 MSVPPEKAEDEAATPGGGEKRPVASTALSAEISLTLLSTAGTATAPLDPAFNKVL 511
| : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 596 ----- 595
| : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 512 MKAVEPKNKADENTPPGSESGAISGVAESSTATLMQSKLMTSLPSWALLTNHFKSTGSF 571
| : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 596 ----- 595
| : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 572 PLPLCARALGASPSSETSLOQLVEKIDROGAVAVYTSAAAGAPTTSAFAPSSASGPNQC 631
| : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 596 -----EC 597
| : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 632 VICLRVLSCPRLRLHAYGQNGGERPFKCKVCGRAFSTRGNLRAHVEYGHKASPARAQNOSC 691
| : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 598 QVCHKRFSSTSNLKTHTLRLHSGEKPYOCKVPAKFTQFVHLKL---HKRLHTRERPHKC 653
| : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 692 PLCOCKFTNAVPTLOOHVBMHLGGQIPNGGTALLPEGGAAQENGSEGSTVSGAGSFPQOQS 751
| : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 654 SOCHKNYIHLCSLKVHLK-----GNCAAPAPGLPLEDTLRIN----- 691
| : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 752 QQPSPEELSEEEDEDEDEEDVTDEDSLARGSESGEKAII--SVRGDSEERASGAEE 809
| : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 692 -----EEIEKFDISDNADRLDEVEDDISYI-----SYVEKEILAVVRKEKEE-TGLK-- 737
| : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 810 VGTVAATAATAGKEMDSNEKTTQOQS-----LPPEPP-----PDSLDQP--QPMF 851
| : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 738 ---VSLQRNMGNGLLSSGCSLYESSDPLMKLPPSNPLPLVPVKVQETVEPMD 788
| : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
```

Search completed: January 13, 2003, 15:20:57
Job time : 53.5336 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 13, 2003, 15:17:40 ; Search time 11.5172 Seconds
(without alignments) updates/sec
3619.260 Million cell

Title: US-09-988-117-1

Perfect score: 5277

Sequence: 1 MAHESERSRLGVPAGEPAE.....PSITSTGLSPRRDDPTIP 1005

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 100 summaries

Database : SWISSPROT_40:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	5167	97.9	1007	1	SAL2_HUMAN
2	4613.5	87.4	1004	1	SAL2_MOUSE
3	1339.5	25.4	1300	1	SAL3_HUMAN
4	1305.5	24.7	1324	1	SAL1_HUMAN
5	1290.5	24.5	1323	1	SAL3_MOUSE
6	1286	24.4	1322	1	SAL1_MOUSE
7	777.5	14.7	1355	1	SALM_DROME
8	745.5	14.1	1402	1	SALM_DROVI
9	376	7.1	1845	1	2236_HUMAN
10	321	6.1	734	1	2N42_HUMAN
11	318	6.0	1839	1	Y296_HUMAN
12	313	5.9	773	1	Z341_HUMAN
13	303	5.7	722	1	2219_HUMAN
14	301.5	5.7	738	1	2N84_HUMAN
15	296.5	5.6	1350	1	XFIN_XENLA
16	293.5	5.6	1191	1	2N91_HUMAN
17	292.5	5.5	982	1	HBL1_CAEBL
18	289.5	5.5	1185	1	DRPL_HUMAN
19	289	5.5	697	1	Y441_HUMAN
20	288	5.5	818	1	KR18_HUMAN
21	280	5.3	1060	1	ZFH1_DROME
22	278.5	5.3	1167	1	2208_HUMAN
23	278	5.3	1163	1	Y222_HUMAN
24	277	5.2	654	1	2133_HUMAN
25	276.5	5.2	947	1	2268_HUMAN
26	276	5.2	1042	1	EV11_MOUSE
27	275	5.2	803	1	2N43_HUMAN
28	271.5	5.1	803	1	2226_HUMAN
29	270	5.1	780	1	ZO84_XENLA
30	269.5	5.1	720	1	PRD4_HUMAN
31	269	5.1	645	1	ZF93_MOUSE
32	269	5.1	682	1	2N45_HUMAN
33	268	5.1	751	1	2184_HUMAN

34	268	5.1	898	1	Z071_XENLA
35	267	5.1	578	1	Z192_HUMAN
36	266	5.0	839	1	ZF95_HUMAN
37	265.5	5.0	739	1	YD49_HUMAN
38	265.5	5.0	1342	1	Z335_HUMAN
39	265	5.0	751	1	Z337_HUMAN
40	265	5.0	1183	1	DRPL_RAT
41	264.5	5.0	1051	1	EV11_HUMAN
42	264	5.0	428	1	ZF26_MOUSE
43	264	5.0	585	1	Z317_HUMAN
44	263	5.0	604	1	GLAS_DROME
45	263	5.0	892	1	HIC1_MOUSE
46	263	5.0	2688	1	HEP1_MOUSE
47	262	5.0	594	1	ZF37_MOUSE
48	262	5.0	616	1	2N93_HUMAN
49	262	5.0	1687	1	Z142_HUMAN
50	260	4.9	595	1	ZN85_HUMAN
51	259.5	4.9	1109	1	TCF8_RAT
52	259	4.9	1411	1	TCOF_HUMAN
53	257.5	4.9	636	1	ZF90_MOUSE
54	257.5	4.9	754	1	2287_HUMAN
55	257	4.9	706	1	Z151_CHICK
56	256.5	4.9	512	1	Z169_HUMAN
57	256.5	4.9	759	1	Z287_MOUSE
58	256	4.9	488	1	Z345_HUMAN
59	256	4.9	717	1	PRDD_HUMAN
60	255	4.8	614	1	ZF28_MOUSE
61	255	4.8	686	1	ZN07_HUMAN
62	255	4.8	733	1	HIC1_HUMAN
63	254	4.8	439	1	ZO28_XENLA
64	253.5	4.8	675	1	ZG20_XENLA
65	253.5	4.8	683	1	Z263_HUMAN
66	253.5	4.8	2717	1	ZEP1_HUMAN
67	253	4.8	670	1	ZN16_HUMAN
68	253	4.8	1117	1	TCF8_MOUSE
69	252.5	4.8	845	1	KRUH_DROME
70	252	4.8	589	1	Z132_HUMAN
71	251	4.7	913	1	Z228_HUMAN
72	250.5	4.7	469	1	Z135_HUMAN
73	250.5	4.7	589	1	ZN44_HUMAN
74	250	4.7	509	1	ZKRI_CHICK
75	248	4.7	582	1	ZF36_HUMAN
76	248	4.7	783	1	ZF25_HUMAN
77	248	4.7	1066	1	Z295_HUMAN
78	247.5	4.7	803	1	ZXDB_HUMAN
79	247	4.7	453	1	Z06_XENLA
80	247	4.7	1403	1	CYCE_DROME
81	246.5	4.7	727	1	CYCE_HUMAN
82	246.5	4.7	1276	1	PRDG_HUMAN
83	245.5	4.7	1029	1	Z197_HUMAN
84	245	4.6	553	1	Z324_HUMAN
85	245	4.6	1267	1	Y211_HUMAN
86	244	4.6	510	1	ZF64_HUMAN
87	244	4.6	794	1	Z151_MOUSE
88	244	4.6	688	1	HRK3_HUMAN
89	243.5	4.6	698	1	Z234_HUMAN
90	242	4.6	821	1	ZN41_HUMAN
91	242	4.6	91	1	SHR1_HUMAN
92	242	4.6	707	1	Z224_HUMAN
93	241.5	4.6	728	1	CYCE_CHICK
94	241	4.6	803	1	Z151_HUMAN
95	241	4.6	2167	1	SHR1_RAT
96	241	4.5	755	1	ZRE1_HUMAN
97	239.5	4.5	489	1	EZT8_HUMAN
98	239	4.5	476	1	Y106_HUMAN
99	238.5	4.5	540	1	Z136_HUMAN
100	238.5	4.5			

ALIGNMENTS

P18751	xenopus lae
O15776	homo sapien
O97218	homo sapien
O92318	homo sapien
O94422	homo sapien
O94258	rattus norv
O03112	homo sapien
P10076	mus musculu
O96966	homo sapien
P13360	drosophila
O91715	mus musculu
O03172	mus musculu
P17141	mus musculu
P35789	homo sapien
P52746	homo sapien
O03923	homo sapien
O62947	rattus norv
O13428	homo sapien
O61967	mus musculu
O98625	gallus gall
O14929	homo sapien
O9e6d9	mus musculu
O14585	homo sapien
O91493	homo sapien
P10078	mus musculu
P17097	homo sapien
O14526	homo sapien
P18747	xenopus lae
O14978	homo sapien
P15822	homo sapien
P17020	homo sapien
O64318	mus musculu
P08135	drosophila
P52740	homo sapien
O9u1j3	homo sapien
P52742	homo sapien
P15621	homo sapien
P30373	gallus gall
P16415	homo sapien
O9u1j5	homo sapien
O9u1j3	homo sapien
P98169	homo sapien
P18749	xenopus lae
O9u1h0	drosophila
P49711	homo sapien
O9ha22	homo sapien
O14709	homo sapien
O5467	homo sapien
O92610	homo sapien
P15622	homo sapien
O60821	mus musculu
O15911	homo sapien
P10074	homo sapien
O14588	homo sapien
P11814	homo sapien
O9y566	homo sapien
O9y213	homo sapien
O08705	gallus gall
O13105	homo sapien
O9wv48	rattus norv
O92766	homo sapien
O9g619	homo sapien
P52737	homo sapien

AL2_HUMAN
D
SAL2_HUMAN STANDARD; PRT; 1007 AA.
O9Y467; O9Y4G1;
15-JUN-2002 (Rel. 41, Created)
15-JUN-2002 (Rel. 41, Last sequence update)
15-JUN-2002 (Rel. 41, Last annotation update)
Sal-like protein 2 (zinc finger protein SAL2) (Hsal2).
SAL2 OR SAL2 OR KIA0360.
Homo sapiens (Human).
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
NCBI_TaxID=9606;
[1]
SEQUENCE FROM N.A. (ISOFORM 1).
MEDLINE=97131507; PubMed=8975705;
Schroeder J., Schu R., Dowe G., Kuehnlein R.P., Jackle H.,
Schroeder B., Schulz-Schaeffer W., Kretschmar H.A., Koehler A.,
Meieler U., Raab-Vetter M., Burkhart E., Engel W., Slick R.;
"Isolation, characterization, and organ-specific expression of two
novel human zinc finger genes related to the Drosophila gene spalt.";
Genomics 38:291-298(1996).
[2]
SEQUENCE OF 11-1007 FROM N.A. (ISOFORM 2).
TISSUE=Brain;
MEDLINE=9734984; PubMed=9205841;
Nagase T., Ishikawa K.-I., Nakajima D., Ohira M., Seki N.,
Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
"Prediction of the coding sequences of unidentified human genes. VII.
The complete sequences of 100 new cDNA clones from brain which can
code for large proteins in vitro.";
DNA Res. 4:141-150(1997).
[3]
SEQUENCE OF 541-1005 FROM N.A. (ISOFORM 2).
Morgan J.W., Ford D., Ma Y., Maizel A.L.;
"Homo sapiens mRNA for zinc finger protein, SAL2 exon 2.";
Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
-1- FUNCTION: Probable transcription factor.
-1- SUBCELLULAR LOCATION: Nuclear (Probable).
-1- ALTERNATIVE PRODUCTS: 2 isoforms; 1 (shown here) and 2; may be
produced by alternative splicing.
-1- TISSUE SPECIFICITY: Highest levels in adult brain (in different
areas). Lower levels in heart; very low levels in kidney and
pancreas.
-1- DEVELOPMENTAL STAGE: In fetal brain exclusively in pontine nuclei.
-1- SIMILARITY: BELONGS TO THE SAL FAMILY OF C2H2-TYPE ZINC-FINGER
PROTEINS.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@sib-sib.ch).

EMBL: X98834; CA667351.1; ALT_INIT.
EMBL: AB002358; BAA21638.1; -;
EMBL: AF465630; BAU74188.1; -;
Genew; HGNC:10526; SAL2.
MIM: 602219; -;
InterPro: IPR000822; Znf.C2H2.
Pfam: PF00096; Zf-C2H2; 7.
PROSITE: PS00028; ZINC_FINGER_C2H2_1; 7.
PROSITE: PS50157; ZINC_FINGER_C2H2_2; 7.
Transcription regulation; DNA-binding; Nuclear protein; Zinc-finger;
Metal-binding; Repeat; Alternative splicing.
ZNF_FING 373 395 C2H2-TYPE.
ZNF_FING 401 423 C2H2-TYPE.
ZNF_FING 631 653 C2H2-TYPE.
ZNF_FING 659 681 C2H2-TYPE.
ZNF_FING 713 733 C2H2-TYPE.
ZNF_FING 911 933 C2H2-TYPE.
ZNF_FING 940 964 C2H2-TYPE.

FT DOMAIN 135 139 POLY-GLY.
FT DOMAIN 158 170 POLY-PRO.
FT DOMAIN 256 265 POLY-SER.
FT DOMAIN 764 775 POLY-GLU.
FT DOMAIN 816 821 POLY-ALA.
FT DOMAIN 838 843 POLY-PRO.
FT VARSPLIC 575 581 LPICARA -> FPVLEP (IN ISOFORM 2).
FT CONFLICT 11 24 OLISDCGFSASEN -> SRLGVPCGEPELIG (IN
REF. 2).
FT CONFLICT 547 547 L -> R (IN REF. 2).
FT CONFLICT 554 554 V -> M (IN REF. 1).
SQ SEQUENCE 1007 AA; 105135 MW; 92D963225A5459E CRC64;

Query Match 97.9%; Score 5167; DB 1; Length 1007;
Best local similarity 99.4%; Pred. No. 7.3e-219;
Matches 984; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 16 GEPALGDASEDEHPQVCAKCAQFTDPTTEFLAHONACSTDPPYAVIIGGEPNNSSA 75
18 GPSASESGDASEDEHPQVCAKCAQFTDPTTEFLAHONACSTDPPYAVIIGGEPNNSSA 77
QY 76 SSEPRPEGHNNPQVMDTEHSNPPDSGSSVPTDPWGPERRGEESGHFLVAATGTAAGG 135
78 SSEPRPEGHNNPQVMDTEHSNPPDSGSSVPTDPWGPERRGEESGHFLVAATGTAAGG 137
QY 136 GGLIASPKIGATPLPEESTPAPPPPPPPPPVSGHLNPLILEELRVLOQRIHOM 195
138 GGLIASPKIGATPLPEESTPAPPPPPPPPPVSGHLNPLILEELRVLOQRIHOM 197
QY 196 QMTEQICQVLLIGSLGQTVGAPASBSELPGTGASTTKPLLPSPKPVOTSKTLASS 255
198 QMTEQICQVLLIGSLGQTVGAPASBSELPGTGASTTKPLLPSPKPVOTSKTLASS 257
QY 256 SSSSSSSGAETPKAFLHYPLHSGHPESAGVGRSHKRTAPSPALPGSTDQILASP 315
258 SSSSSSSGAETPKAFLHYPLHSGHPESAGVGRSHKRTAPSPALPGSTDQILASP 317
QY 316 HLAPESTGLLAOCLGAARGLAETASPGLLKPKNGSELGYGEVWGPLEKGGRRKCRF 375
318 HLAPESTGLLAOCLGAARGLAETASPGLLKPKNGSELGYGEVWGPLEKGGRRKCRF 377
QY 376 CAKYFGSALQILHRSHTGERPKYCNVCGNRFTRGNLKAHFNHREKRPVQNMHPY 435
378 CAKYFGSALQILHRSHTGERPKYCNVCGNRFTRGNLKAHFNHREKRPVQNMHPY 437
QY 436 PEHLDYVITSSGLPYGMSVPEPEKAEAEATPGGVERKPLVASTALSTESLTILSTSA 495
438 PEHLDYVITSSGLPYGMSVPEPEKAEAEATPGGVERKPLVASTALSTESLTILSTSA 497
QY 496 GTATAPGILPAFNFVLMKAVEPKKADENTPPGSESAISGVAESSTATLMOLSKLMTSL 555
498 GTATAPGILPAFNFVLMKAVEPKKADENTPPGSESAISGVAESSTATLMOLSKLMTSL 557
QY 556 PSMALLTNHFKSTGSEPLPLCARALGASPSETSKLOQLVEKIDRGAVAVSAASAPPT 615
558 PSMALLTNHFKSTGSEPLPLCARALGASPSETSKLOQLVEKIDRGAVAVSAASAPPT 617
QY 616 SAPAPSSASSGPNOCVLCRLVSCPRALFLHYGOHGERFCKVCVCGRAFTSGNLRAH 675
618 SAPAPSSASSGPNOCVLCRLVSCPRALFLHYGOHGERFCKVCVCGRAFTSGNLRAH 677
QY 676 FVGHKASPAARAQNSCPICQKFTNAVTLOQHVHMLGGQIPNGGTALPEGGAAQENGS 735
678 FVGHKASPAARAQNSCPICQKFTNAVTLOQHVHMLGGQIPNGGTALPEGGAAQENGS 737
QY 736 EOSTVSGAGSPPOOQSOQPSPEEELSEEEDEDEDEEDVDDESLARGSSGGEKALS 795
738 EOSTVSGAGSPPOOQSOQPSPEEELSEEEDEDEDEEDVDDESLARGSSGGEKALS 797
QY 796 VRGDSSEASGAEEVGTAAATAGKEMDSNEKTTQOOSLPPPPPPDLDPQPPDEGSS 855
798 VRGDSSEASGAEEVGTAAATAGKEMDSNEKTTQOOSLPPPPPPDLDPQPPDEGSS 857

Y	856	GVLGKKEGGKPERSSSPASALTPGEATSVTLVEELSLQEAAMRKPEGSSSRKACEVCG	915
Y	858	GVLGKKEGGKPERSSSPASALTPGEATSVTLVEELSLQEAAMRKPEGSSSRKACEVCG	917
Y	916	QAFPSQAALAEHOKTHPEKPLFTVCFCROGFLERATLKHHMLAHNOVPAPHPGPONT	975
Y	918	QAFPSQAALAEHOKTHPEKPLFTVCFCROGFLERATLKHHMLAHNOVPAPHPGPONT	977
Y	976	AALSLVPGCSPSTSTGSLSPFRKDDPTIP	1005
Y	978	AALSLVPGCSPSTSTGSLSPFRKDDPTIP	1007
D	RESULT 2		
D	SAL2_MOUSE	STANDARD;	PRT; 1004 AA.
D	SAL2_MOUSE		
C	Q90X96;		
T	15-JUN-2002 (Rel. 41, Created)		
T	15-JUN-2002 (Rel. 41, Last sequence update)		
T	15-JUN-2002 (Rel. 41, Last annotation update)		
T	Sal1-like protein 2 (Spalt-like protein 2) (Msal-2).		
T	SAL2 OR SAL2.		
E	Mus musculus (Mouse).		
C	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
C	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
X	NCBI_TaxId=10090;		
X	[1]		
P	SEQUENCE FROM N.A.		
P	TISSUE=Brain;		
P	MEDLINE=20069074; PubMed=10602995;		
A	Kohlhaase J., Altman M., Archangelo L., Dixkens C., Engel W.;		
A	"Genomic cloning, chromosomal mapping, and expression analysis of		
T	Msal-2."		
T	Mamm. Genome 11:64-68(2000).		
L	-1- FUNCTION: Probable transcription factor.		
L	-1- SUBCELLULAR LOCATION: Nuclear (Probable).		
L	-1- TISSUE SPECIFICITY: Expressed throughout embryonic development. In		
L	adult predominantly in brain.		
L	-1- SIMILARITY: BELONGS TO THE SAL FAMILY OF C2H2-TYPE ZINC-FINGER		
L	PROTEINS.		
C	-----		
C	This SWISS-PROT entry is copyright. It is produced through a collaboration		
C	between the Swiss Institute of Bioinformatics and the EMBL outstation -		
C	the European Bioinformatics Institute. There are no restrictions on its		
C	use by non-profit institutions as long as its content is in no way		
C	modified and this statement is not removed. Usage by and for commercial		
C	entities requires a license agreement (See http://www.isb-sib.ch/announce/		
C	or send an email to license@isb-sib.ch).		
C	-----		
C	EMBL: AJ007396; CAB65274.1; -.		
C	MGD: MGI:1354373; Sal12.		
C	InterPro: IPR000822; Znf_C2H2.		
C	PIfam: PF00096; zf-C2H2; 7.		
C	SMART: SM00355; Znf_C2H2; 7.		
C	PROSITE: PS00028; ZINC_FINGER_C2H2_1; 7.		
C	PROSITE: PS00157; ZINC_FINGER_C2H2_2; 7.		
W	Transcription regulation; DNA-binding; Nuclear protein; Zinc-finger;		
W	Metal-binding; Repeat.		
W	ZN_FING	372	394
W	ZN_FING	400	422
W	ZN_FING	629	651
W	ZN_FING	657	679
W	ZN_FING	689	711
W	ZN_FING	908	930
W	ZN_FING	937	961
W	DOMAIN	135	139
W	DOMAIN	158	170
W	DOMAIN	256	264
W	DOMAIN	762	772
W	DOMAIN	834	840
W	SEQUENCE	1004 AA;	104975 MW; E46DD2C9C26A4C92 CRC64;

Query Match 87.4%; Score 4613.5; DB 1; Length 1004;

Best Local Similarity 89.0%; Pred. No. 1e-194;			
Matches 881; Conservative 25; Mismatches 81; Indels 3; Gaps 3;			
QY	16	GEPAELGGDASEEDHPQVACCKCAQFTDPTFEFLAQNNACSTPPVAVIIGGQENPNNSA	75
DB	18	GPSASENGDAEEHHPQVACCKCAQFSDPTFEFLAQNNACSTPPVAVIIGGQENPNNSA	77
QY	76	SSEPREGHNNPOVMDEHNSPPDSGVPTDPTMGPERGEGSGHFLVAATGTAAGC	135
DB	78	SSAPREGHNSQVMDEHNSPPDSGVPTDPTMGPERGEGSGHFLVAATGTAAGC	137
QY	136	GGILLASPKLGATPLPESTPAPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPP	195
DB	138	GGILLASPKLGATPLPESTPAPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPP	197
QY	196	QMEQICROYLVLLGSLQOTVAPASPSLPGTGATSSPTPLPLPSPIKPVQTSKLASS	255
DB	198	QMEQICROYLVLLGSLQOTVAPASPSKLPCTGAASSTKPLPLPSPIKPVQTKTLA-S	256
QY	256	SSSSSSSGAETPKQAFPHLYHPLGSOHPFSAGVGRSHKPTPAPAPALPGSTDOLIASP	315
DB	257	SSSSSSSGAETPKQAFPHLYHPLGSOHPFSAGVGRSHKPTPAPAPALPGSTDOLIASP	316
QY	316	HLAFPTTGLLAOCGLAARGLEATASPGLLKPKNGSGELSTGEVMOPLKPGGRHKCF	375
DB	317	HLAFPTTGLLAOCGLAARGLEAASPGLLKPKNGSGELSTGEVMOPLKPGGRHKCF	376
QY	376	CARVFGSDSALQTHLRSHITGERPKYCNVGNFTTRGNLKVHHRREKPHYOMNHPV	435
DB	377	CARVFGSDSALQTHLRSHITGERPKYCNVGNFTTRGNLKVHHRREKPHYOMNHPV	436
QY	436	PEHLDVYITSSGFLPYGMSVPEPEKAEAEATPGGVERKPLVASTATLSTESLTLTSTA	495
DB	437	PEHLDVYITSSGFLPYGMSVPEPEKAEAEATPGGVERKPLVASTATLSTESLTLTSTG	496
QY	496	GTAATPGLPAFNKFLVLMKAVEPKNKADENTPGSEGSATISGVAESSTATLMQSKMTSL	555
DB	497	STAVAPGLPTFNKFLVLMKAVEPKSKADENTPGSEGSATISGVAESSTATLMQSKMTSL	556
QY	556	PSNALLTNHFKSTGSPFLPLCARALGASPSMTSKLOOLYEKIDROGAVVATSAAGAPT	615
DB	557	PSNALLTNHFKSTGSPFPVYLEPLGASPSMTSKLOOLYEKIDROGAVVATSAAGAPT	616
QY	616	SAPAPSSSASSGPNQVCYCLRVLSCPRALRLHYGOGGERPKCYGGRAFSTRGNLRAH	675
DB	617	SAPAPSSSA-SGPNQVCYCLRVLSCPRALRLHYGOGGERPKCYGGRAFSTRGNLRAH	675
QY	676	FVGHKASPARAONSCPTICQKKFTNAVTLQOHVRLHGGQIPNGCTALPBGCGAAQENG	735
DB	676	FVGHKTSPPARAONSCPTICQKKFTNAVTLQOHVRLHGGQIPNGGSALSBEGGAQENG	735
QY	736	EOSTVSGAGSPFOQSOQSPPEELSEEEDEDEEDVTDSDSLAGRSGSGEKRAIS	795
DB	736	EOSTASGPGSPFOQSOQSPPEEMS-EEEDDEDEEDVTDSDSLAGRSGSGEKRAIS	794
QY	796	VRGDSSEASGAEEEGYTAATAATAGKEMDSNEKTTQOSSLPPEPPPSLDQCPQMEQSS	855
DB	795	VRGDSSEASGAEEEGYTAATAATYKEMDSNEKAPHTLPPPEPPPNLDHPQMEQSTS	854
QY	856	GVLGKKEEGKPERSSSPASALTPEGEATSVTLVEELSLQEAAMRKPEGSSSRKACEVCG	915
DB	855	DVSGAMEEAEALTEGSSPMALTOEGEGTSTPLVEELINPEAMKKDGESSGRKACEVCG	914
QY	916	QAFPSQAALAEHOKTHPKREGPLFTVCFCROGFLEKATLKHHMLAHNOVPAPHPGPONT	975
DB	915	QSFPTQALAEHOKTHPKREGPLFTVCFCROGFLDKATLKHHMLAHNOVPAPHPGPONT	974
QY	976	AALSLVPGCSPSTSTGSLSPFRKDDPTIP	1005
DB	975	ATLSLVPGCSSSIPSPGLSPFRKDDPTMP	1004
RESULT 3			

SAL3_HUMAN STANDARD; PRT; 1300 AA.
ID SAL3_HUMAN Q9BGH1:
15-JUN-2002 (Rel. 41, Created)
15-JUN-2002 (Rel. 41, Last sequence update)
15-JUN-2002 (Rel. 41, Last annotation update)
Sal-1-like protein 3 (zinc finger protein SAL3) (hsSAL3).
SAL3.
Homo sapiens (Human).
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
NCBI_TaxID=9606;
[1]
SEQUENCE FROM N.A. (ISOFORM 3).
3A Gooding R., Angelicheva D., Blechschmidt K., Snoboda K., Molnar M.,
3B Tourneir I., Kalaydjieva L.,
3C "Exclusion of HSAL3 and refinement of the region for the CCFDN
3D gene."
3E Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
3F [2]
3G SEQUENCE OF 29-1300 FROM N.A., AND ALTERNATIVE SPLICING.
3H MEDLINE=20079154; Pubmed=10610715;
3I Kohlbase J., Hausmann S., Stojmenovic G., Dixkens C., Bink K.,
3J Schulz-Schaeffer W., Altman M., Engel W.,
3K "SAL3, a new member of the human spalt-like gene family, maps to
3L 18q23."
3M Genomics 62:216-222(1999).
3N -1- FUNCTION: Probable transcription factor.
3O -1- SUBCELLULAR LOCATION: Nuclear (Probable).
3P -1- ALTERNATIVE PRODUCTS: 4 isoforms; 1, 2, 3 (shown here) and 4, are
3Q produced by alternative splicing. Isoforms 1 and 2 lack two zinc
3R finger domains and are the major isoforms.
3S -1- TISSUE SPECIFICITY: Widely expressed in adult with highest levels
3T in heart. Expressed in fetal brain (in neurons of hippocampus,
3U cerebellum and brainstem).
3V -1- DEVELOPMENTAL STAGE: In fetal brain of the 24th gestational week.
3W -1- SIMILARITY: BELONGS TO THE SAL FAMILY OF C2H2-TYPE ZINC-FINGER
3X PROTEINS.
3Y -----
3Z This SWISS-PROT entry is copyright. It is produced through a collaboration
40 between the Swiss Institute of Bioinformatics and the EMBL outstation-
41 the European Bioinformatics Institute. There are no restrictions on its
42 use by non-profit institutions as long as its content is in no way
43 modified and this statement is not removed. Usage by and for commercial
44 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
45 or send an email to license@sib-sib.ch).
46 -----
47 EMBL; AF347021; AAK18311.1; -;
48 EMBL; AJ007421; CAB65124.1; -;
49 HSSP; P07248; IARE.
50 Genew; HGNC:10527; SAL3.
51 MIM; 605079; -;
52 InterPro: IPR000822; znf.C2H2.
53 Pfam: PF00096; zf.C2H2; 10.
54 SMART: SM00355; znf.C2H2; 10.
55 PROSITE; PS00028; ZINC_FINGER_C2H2_1; 9.
56 PROSITE; PS00157; ZINC_FINGER_C2H2_2; 8.
57 Transcription regulation; DNA-binding; Nuclear protein; Zinc-finger;
58 Metal-binding; Repeat; Alternative splicing.
59 ZN_FING 420 442 C2H2-TYPE.
60 ZN_FING 448 470 C2H2-TYPE.
61 ZN_FING 679 701 C2H2-TYPE.
62 ZN_FING 707 729 C2H2-TYPE.
63 ZN_FING 739 761 C2H2-TYPE.
64 ZN_FING 977 999 C2H2-TYPE.
65 ZN_FING 1005 1027 C2H2-TYPE.
66 ZN_FING 1113 1135 C2H2-TYPE.
67 ZN_FING 1141 1163 C2H2-TYPE.
68 DOMAIN 198 201 POLY-ALA.
69 DOMAIN 213 216 POLY-GLN.
70 DOMAIN 897 902 POLY-SER.
71 VARSPPLIC 1 133 MISSING (IN ISOFORM 2 AND ISOFORM 4).
72

FT VARSPPLIC 973 1044 MISSING (IN ISOFORM 1 AND ISOFORM 2).
FT CONFLICT 235 235
FT CONFLICT 593 593
FT CONFLICT 787 787
FT CONFLICT 797 802
FT CONFLICT 808 808
FT CONFLICT 1138 1138
FT CONFLICT 1141 1141
SQ SEQUENCE 1300 AA; 135371 MW; 82CF3BDCB6D59150 CRC64;
Query Match 25.4%; Score 1339.5; DB 1; Length 1300;
Best local Similarity 33.2%; Pred. No. 9.6e-52;
Matches 385; Conservative 115; Mismatches 343; Indels 317; Gaps 43;
QY 16 GEPAEELGGDASEE-----DHPVYCAKCCAGQFDPTPEFLANQNCSTDPFVWYITGGQENPN 71
DB 31 GGEAEADSGPEPSRSGEETSVCEKCAEFKWAQFLEHQRCSTKPLPVYLV----- 82
QY 72 NSSASSEPREGHNNPQVMDENHSNPDSSSVPTDPTGPERGEGSGHFLVAATCTA 131
DB 83 HEDAPAP-----HEDFE-----PSPASSPERASE-----AAEEAG 116
QY 132 AGGGGGLILASPK---LGATLPPESTPAPPPPPPPGVSGHLN----- 176
DB 117 AEGAEAEARVEKEAEEDPADGTRAPRPPAPAPPTPAGABSTVTLBALLSTKV 176
QY 177 -----PLLEELRYLQCGQHOMQMTEDICRYVLLG-- 209
DB 177 AVAOFSGARAGSGAGGVAANAVALPILDLQMALQDQIHQLDLEQIRSOVALMQRP 236
QY 210 -----SLGQTVGAPA-SPSELPGTASTSTKPLPLPSPIKPVQTSKTLASSSSSS 262
DB 237 PPRPRLSPAAAP 277
QY 263 SGAEPTKQAFILYHPL-----GSQHPPSAGCVGSHKP-----TPAPSPALPGSTDLT 312
DB 278 SGPAAP--AAFEAGAPLSRESGASTP---GGPAPPSAPAPSAAPAPAPAPAPAPAP 332
QY 313 ASPHLAPFSTGLLA-AQCGAARGLEA-----TPASGLKP----- 348
DB 333 ASSQPOSASTPPALAPPSLLGAAPGLSPPLPOTSASGVTFPNPLVSIATANALDPLSA 392
QY 349 --KNGSGE--LSYGEVWGLEPKGGRKRCRCACAVFGSDSALQIHLRSHTGRRPYCNV 403
DB 393 LKMRKRGKRPVAVSEVERKASAEDEPFKHKRCACAVFGSDSALQIHLRSHTGRRPKCN 452
QY 404 CGNRFTTRGALKYHFNHNRKRYPHVQANRPHRYPEHLDYITSSGLPYGKSVPEKAEESA 463
DB 453 CGNRFSTKGNLKVHFNHNRKRYPHIOMNRPVPEYLDVNPVTSGLPYGMSLPE----- 506
QY 464 ATPGGGVERRPLVASTATLSTESLTLSAGTATAPGLPAFNKFLMKAVPPKNKADE 523
DB 507 -----KPV---ITWLDKRPVLPVPTVSVGLDLPPTVPAHAGVADSPSATPAPASRPQ 534
QY 524 NTPPGSE-----GSAISGVAESSTATL----- 545
DB 555 RSPASSECAASLSPGLNHVSGVATGAESPQSLGPGVPTKAPVSLPCTNARAGDAPVG 614
QY 546 MQLSKMTSL-----PSMALLTNHFKSTGSPRLDLCANALASPSSETKLDQVLEK 596
DB 615 AASAAPATVDGAPSTLSGSGPLPAVVSQFKA--QPPFGGLDMS--QVSETSKLDQVLEN 670
QY 597 IDROGAAVATSAASGAPTTAPAPASSASSSPNOCVICLRVYLSCPREALYHGOHGERP 656
DB 671 IDKK-----MDPNQCVICHRVYLSGCSALKMHRRTHTGERP 706
QY 657 FCKYCGRAPESTRGNLRHFVGRKASPAARAQNSPICQKKTAAVTLQOHVHMLGGQI 716
DB 707 FCKYICGRAFTTKGNLKHGVRHAKRPPRLRVQHSCLPICQKKTAAVTLQOHIMHGGQI 766
QY 717 PRCGTALPEGGGAA-----QENGSGQST-----VS 741
DB 767 PN--TPLPGEFODAMDSELAYDCKNAETLSSYDDMDENSMEDDAELKDAATDAPKPLLS 824


```
Oy 742 GAGTTPQOQSQPSPPEEELSEEEDEDEDED-VTDEDLQAG-RGSESGEKAIVRGD 799
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 825 YAGSCP-----PSPSPVSISSIALEENQMKIDVMSCCOQLTGLKSVENGSGESDRLSND 878
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Oy 800 SEEAAGAEVEYGVAAATATGKEMDSNEKTTQOSSLPPPPPPDSLDQPMEGSGSSGVLG 859
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 879 SSSAAGDLE-----SRSGSGPALVS-ESSSSQALSPAPNCGESRSKSP-----GLG 923
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Oy 860 GKEEGS-----KPERSSSPASALTPEGEATSVTLVEELSLQEA-MRKEPSSSRKACEVC 914
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 924 APEQEIEILKTERPDSPAPAGSGAGBRAGIKIEAEVPSLLFLSREKGCPS-TVCGVC 982
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Oy 915 GCAFPSQALAEHQKHPKCGPLFTYCFRCROGFLERATLKKHML---LAHQVQFAPRH- 970
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 983 GPFCKSALEIYRSHTKERP-FVCAICRCGCGSTMGNLKHLLTHRLKELPSQLFDPNF 1041
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Oy 971 --GPQINIALSLVPGCGSPSI 988
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 1042 AICPSSO-STPSSLSSAAPTM 1060

RESULT 4
ID   SALL_HUMAN
AC   Q9NSC2; Q9NSC3; Q99881; Q9P1R0;          PRT;   1324 AA.
DT   15-JUN-2002 (Rel. 41, Created)
DT   15-JUN-2002 (Rel. 41, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Sall-like protein 1 (zinc finger protein SALL1) (Spalt-like
DE   transcription factor 1) (Hsall1).
GN   SALL1 OR SALL.
OS   Homo sapiens (human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A., DISEASE, AND VARIANTS SER-150 DEL; SER-150 INS AND
RP   GLY-159.
RX   MEDLINE=99138656; PubMed=9973281.
RA   Kohlhaase J., Taschner P.E.M., Burfelde P., Pasche B., Newman B.,
RA   Blanck C., Breuning M.H., ten Kate L.P., Maaswinkel-Mooy P.,
RA   Mittella B., Seidel J., Kirkpatrick S.J., Pauli R.M., Margowski D.S.,
RA   Devriendt K., Proesmans W., Gabrielli O., Coppa G.V.,
RA   Wesely-van Swaay E., Trembath R.C., Schinzel A.A., Reardon W.,
RA   Seemanova E., Engel W.,
RA   "Molecular analysis of SALL1 mutations in Townes-Brocks Syndrome.",
RA   Am. J. Hum. Genet. 64:435-445(1999).
RL   [2]
RN   [2]
RP   SEQUENCE OF 1-26 FROM N.A., DISEASE, AND VARIANTS SER-164 DEL AND
RP   GLU-1265.
RX   MEDLINE=20004537; PubMed=10533063.
RA   Marlin S., Blanchard S., Lacombe D., Denoyelle F., Alessandri J.-L.,
RA   Calciolari E., Drouin-Garraud V., Ferraz F.G., Fourminteraux A.,
RA   Philip N., Toubanc J.E., Petit C.,
RA   "Townes-Brocks syndrome: detection of a SALL1 mutation hot spot and
RA   evidence for a position effect in one patient.",
RA   Hum. Mutat. 14:377-386(1999).
RL   [3]
RN   [3]
RP   SEQUENCE OF 26-1324 FROM N.A.
RX   MEDLINE=97131507; PubMed=8975705;
RA   Kohlhaase J., Schuh R., Dove G., Kuehnlein R.P., Jaeckle H.,
RA   Schroeder B., Schuiz-Schaeffer W., Kretzschmar H.A., Koehler A.,
RA   Mueller U., Raab-Vetter M., Burkhardt E., Engel W., Stick R.;
RA   "Isolation, characterization, and organ-specific expression of two
RA   novel human zinc finger genes related to the Drosophila gene spalt.",
RA   Genomics 38:291-298(1996).
RL   [4]
RN   [4]
RP   SEQUENCE OF 313-345 FROM N.A., AND DISEASE.
RX   MEDLINE=98085876; PubMed=9425907;
RA   Kohlhaase J., Witschmann A., Reichenbach H., Froster U., Engel W.,
RA   "Mutations in the SALL1 putative transcription factor gene cause
RA   Townes-Brocks syndrome.",
```

```
RL Nat. Genet. 18:82-83(1998).
RN [5]
RX DISEASE.
RP MEDLINE=20381976; PubMed=10928856;
RA Engels S., Kohlhaase J., McGaughran J.;
RT "A SALL1 mutation causes a branchio-oto-renal syndrome-like
RT phenotype.";
RL J. Med. Genet. 37:458-460(2000).
CC -I- FUNCTION: transcriptional repressor involved in organogenesis (by
CC similarity).
CC -I- SUBUNIT: interacts with HDAC1, HDAC2, RBBP4, RBBP7, MTA1 and MTA2
CC (By similarity). Probably associates with NuRD histone deacetylase
CC complex (HDAC).
CC -I- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC -I- TISSUE SPECIFICITY: Highest levels in kidney. Lower levels in
CC adult brain (enriched in corpus callosum, lower expression in
CC substantia nigra) and liver.
CC -I- DEVELOPMENTAL STAGE: in fetal brain exclusively in neurons of the
CC subependymal region of hypothalamus lateral to the third
CC ventricle.
CC -I- DISEASE: Defects in SALL1 are the cause of Townes-Brocks syndrome
CC (TBS); a rare, autosomal dominant malformation syndrome with a
CC combination of imperforate anus, triphalangial and supernumerary
CC thumbs, malformed ears and sensorineural hearing loss.
CC -I- DISEASE: Defects in SALL1 are in rare cases associated with
CC hemifacial microsomia (HFM) or Goldenhar syndrome; a common birth
CC defect involving first and second branchial arch derivatives. The
CC highly variable phenotype shows in addition to craniofacial
CC anomalies cardiac, vertebral, and central nervous system defects.
CC -I- DISEASE: Defects in SALL1 are associated with branchio-oto-renal
CC (BOR) syndrome, an autosomal dominant disorder manifested by
CC various combinations of preauricular pits, branchial fistulae or
CC cysts, lacrimal duct stenosis, hearing loss, structural defects of
CC the outer, middle, or inner ear, and renal dysplasia. Associated
CC defects include asymmetric habitus, long narrow facies, constricted
CC palate, deep overbite, and myopia. Hearing loss may be due to
CC Mondini type cochlear defect and stapes fixation.
CC -I- SIMILARITY: BELONGS TO THE SALL FAMILY OF C2H2-TYPE ZINC-FINGER
CC PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; Y18265; CAB41400.1; -.
DR EMBL; Y18264; CAB41399.1; -.
DR EMBL; X98833; CAB41399.1; JOINED.
DR EMBL; AF017655; AAB99908.1; -.
DR EMBL; AF074949; AAF19263.1; -.
DR HSSP; P15822; IBBP.
DR Genew; HGNC:10524; SALL1.
DR MIM; 602218; -.
DR MIM; 107480; -.
DR MIM; 164210; -.
DR MIM; 113650; -.
DR InterPro; IPR000822; ZnF_C2H2.
DR Pfam; PF00096; ZF-C2H2; 10.
DR ProDom; PD000003; ZnF-C2H2; 1.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 9.
DR PROSITE; PS50157; ZINC_FINGER_C2H2_2; 9.
KW Transcription regulation; Repressor; DNA-binding; Nuclear protein;
KW Zinc-finger; Metal-binding; Repeat; Polymorphism.
FT ZN_FING 449 471 C2H2-TYPE.
FT ZN_FING 477 499 C2H2-TYPE.
FT ZN_FING 706 728 C2H2-TYPE.
FT ZN_FING 734 756 C2H2-TYPE.
FT ZN_FING 766 788 C2H2-TYPE.
FT ZN_FING 1001 1023 C2H2-TYPE.
FT ZN_FING 1029 1051 C2H2-TYPE.
```



```

CC      This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC
CC      EMBL: X97581; CAA66196.1; -.
CC      DR      HSSP: P07248; IARE.
CC      DR      MGD; MGI:109285; Sal13.
CC      DR      InterPro: IPR000822; Znf_C2H2.
CC      DR      Pfam: PF00096; Zf-C2H2; 10.
CC      DR      SMART: SM00355; Znf_C2H2; 9.
CC      DR      PROSITE: PS00028; ZINC_FINGER_C2H2_1; 9.
CC      DR      PROSITE: PS00157; ZINC_FINGER_C2H2_2; 8.
CC      KM      Transcription regulation; DNA-binding; Nuclear protein; Zinc-finger;
CC      Metal-binding; Repeat; Alternative splicing.
CC      FT      NON_TER 1
CC      FT      ZN_FING 398 420 C2H2-TYPE.
CC      FT      ZN_FING 426 448 C2H2-TYPE.
CC      FT      ZN_FING 663 685 C2H2-TYPE.
CC      FT      ZN_FING 691 713 C2H2-TYPE.
CC      FT      ZN_FING 723 745 C2H2-TYPE.
CC      FT      ZN_FING 968 990 C2H2-TYPE.
CC      FT      ZN_FING 996 1018 C2H2-TYPE.
CC      FT      ZN_FING 1104 1126 C2H2-TYPE.
CC      FT      ZN_FING 1132 1154 C2H2-TYPE.
CC      FT      DOMAIN 119 123 POLY-PRO.
CC      FT      DOMAIN 188 191 POLY-GLN.
CC      FT      DOMAIN 345 348 POLY-SER.
CC      FT      DOMAIN 881 886 POLY-SER.
CC      FT      VARSPIC 964 1035 MISSING (IN ISOFORM 2).
CC      SO      SEQUENCE 1323 AA; 139070 MW; 86D83EDEAF2CEDD CRC64;
CC
CC      Query Match 24.5%; Score 1290.5; DB 1; Length 1323;
CC      Best Local Similarity 31.8%; Pred. No. 1.3e-49;
CC      Matches 373; Conservative 137; Mismatches 343; Indels 319; Gaps
CC
CC      QY      17 EPAELGDA-SEEDHPQVCAKCAQFTDPTERLHAONACSTPRTVWVYIIGQENPNNSA 75
CC      Db      6 EDADSGESRSGESETSVCCKCAEFKAWADPLQHKKTCTKNPLVIV-----H 54
CC
CC      QY      76 SEEPREGHNHPQVMDTEHSNPPDSGVPPD---PTWPERRGESSGHFLVATGTAA 132
CC      Db      55 DDEPAPSEDFPE--PSPASSPDRTESEVAEVAETGESE-----VKATKES 102
CC
CC      QY      133 GGGGGLILASPKLGATPLPESTAPRPPPPPPPP----- 168
CC      Db      103 MD-----VEVSTDGPR---PGSPVPPPPALPQPEREASMPETNTVLETLSTKVAVAQ 155
CC
CC      QY      169 -----GWGS--GHLNPILEELRYLQROIHQMOMTEOICROYLLGLSGQTV 215
CC      Db      156 FSGAGAGCTTGAGGSVGAIVAPMTEQLVALQOOQIHOLQLEIGIRSYV---GLMSROP 212
CC
CC      QY      216 GAPASSEL-----PGCTASTKRLPLFPSP---IKPQOT 248
CC      Db      213 GPLPKESASAPRGNTSVQLGLTPHAAQLLSAGPATAVASGSSTLPAPFDPQHLISQPASG 272
CC
CC      QY      249 SKTLAASSSSSSSSGAEPTKQAEFFHLNPLISQNP-----FSAGGVGRSHKP---TP 298
CC      Db      273 TSPICSTSAAPDSGA-----HPACSTGAPAPGAVAASSTVGNNAVQPNASTP 321
CC
CC      QY      299 -----APSPALPGSTDQLIASPHLAPRSTGLAAOCLGARGLEATASPG 344
CC      Db      322 PALGPGLTSSASMLPNPLPQT-----SSSIVFPNPLVSIAA---TANALPLSA-- 370
CC
CC      QY      345 LKPKNGS-GELSYGEVWGPLEKPGRCRCACKFGSSALQIHLRSHSGEPYKCNV 403
CC      Db      371 LKHHRGKPPNVSVFERKASADEPFTKHCRFCACAKFGSSALQIHLRSHSGEPYKCNV 430
CC      404 CGNRFTTRGMLKVFHHRHREKRYPHVQMPNHPAREHLDYVTSSTGLPQGMASVPEK----- 458

```

[illegible]

Copeland N.G., Gilbert D.J., Jenkins N.A., Scully S., Lacey D.L., Katsuki M., Aashima M., Yokota T.; "Murine homolog of Snail1 is essential for ureteric bud invasion in kidney development." Development 128:3105-3115(2001).

[3]

FUNCTION, AND ASSOCIATION WITH HDAC.

Medline=21964044; PubMed=11836251;

Kiefer S.M., McDill B.W., Yang J., Rauchman M.; "Murine snail1 represses transcription by recruiting a histone deacetylase complex." J. Biol. Chem. 277:14869-14876(2002).

-1- FUNCTION: Transcriptional repressor involved in organogenesis. Homozygous deletion of Snail1 results in an incomplete ureteric bud outgrowth, a failure of tubule formation in the mesenchyme and an apoptosis of the mesenchyme.

-1- SUBUNIT: Interacts with HDAC1, HDAC2, RBBP4, RBBP7, MTA1 and MTA2. Probably associates with NURD histone deacetylase complex (HDAC).

-1- SUBCELLULAR LOCATION: Nuclear.

-1- TISSUE SPECIFICITY: Expressed in the metanephric mesenchyme surrounding ureteric bud.

-1- SIMILARITY: BELONGS TO THE SNAIL FAMILY OF C2H2-TYPE ZINC-FINGER PROTEINS.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@sib-sib.ch).

EMBL: AJ271914; CAC09602.1; -

EMBL: AJ271915; CAC09602.1; JOINED.

EMBL: AB051409; BAB5673.1; -

HSSP: P18822; 1BBO.

MCD, MGI:1889585; Snail1.

InterPro: IPR000822; Znf_C2H2.

Pfam: PF00096; Zf_C2H2; 10.

ProDom: PD000003; Znf_C2H2; 1.

SMART: SM00355; Znf_C2H2; 9.

PROSITE: PS00028; ZINC_FINGER_C2H2_1; 9.

PROSITE: PS50157; ZINC_FINGER_C2H2_2; 9.

Transcription regulation; Repressor; DNA-binding; Nuclear protein; Zinc-finger; Metal-binding; Repeat.

2N_FING 450 472 C2H2-TYPE.

2N_FING 478 500 C2H2-TYPE.

2N_FING 705 727 C2H2-TYPE.

2N_FING 733 755 C2H2-TYPE.

2N_FING 765 787 C2H2-TYPE.

2N_FING 1000 1022 C2H2-TYPE.

2N_FING 1028 1050 C2H2-TYPE.

2N_FING 1133 1155 C2H2-TYPE.

2N_FING 1161 1183 C2H2-TYPE.

2N_FING 133 139 POLY-THR.

DOMAIN 237 240 POLY-SER.

DOMAIN 1143 1146 POLY-SER.

CONFLICT 165 165 C -> S (IN REF. 2).

CONFLICT 167 167 S -> T (IN REF. 2).

CONFLICT 1271 1271 S -> SS (IN REF. 2).

SEQUENCE 1322 AA: 140229 MW: 24477 FE0895F5C CRC64;

Query Match 24.4%; Score 1286; DB 1; Length 1322;

Best Local Similarity 32.1%; Pred. No. 2,1e-49;

Matches 383; Conservative 145; Mismatches 340; Indels 326; Gaps 50;

16 GEPRLGDAEEDHPQVAKCAQPTDPEFLAHQNAQCSDDPVMVITGGQENPNSSA 75

31 GQPSR--PTKSKDAH--VGRCCAEFFELSDLLAKKSCCTKNQVLVI---NESPA 83

76 SSEPRRGHNHPQVADTEHNSNPDDSSSVPTDPTGPERGSESSGHFLVAATGT 130

Db 84 TEPGPR-SLNDPDDQMKDANKADQEDCSLSEPKGLDR--EESMEVEVPATTTTTT 140

QY 131 AAGGGGLILASPKLCAATP-----LPP----- 152

Db 141 GSGGGGSLTSGVTNTTPSCHSGCSTGSAITTSLPOLGDLTTLGNFSVINSVIENT 200

QY 153 ESTP-APPPPPPPPPGSGSHINPLILELRVLAQOROHOMOMTEICROVLLGSL 211

Db 201 QSTKVAVAFQSGARCGAGSGKLLISTLMQALALQOQIHQDLLEIRHILLAS- 259

QY 212 GQTVGAPASPSELPGTGT-ASSTRKPLLPSPKIR-----PVQSKTLASS- 256

Db 260 -QADLPAPAP--IPSGGTLRTSANPLTTLSSHLTSQLAVAAQSLASIASANISGVKOL 317

QY 257 -----SSSSS-----SGAETKQAFHLVHPPLGSOHPFAGVGRSKPPTA-----PS 301

Db 318 PHVOLPQSSSGTSIYPPSGGTSFPMSTIVTAAPVPPSSKV-ASNAGASHVSSPAVASSS 376

QY 302 PA-----LPGSTDLIASP---HLAFTTGLLA--AQCLGAAGLEATASPLGPKN 350

Db 377 PAFALISLSPSSNPPLPQPTPANAVPPTPLPNIAITTEDLNS---LSALAQRRKRPN 433

QY 351 GSGELSTGEVMPLEKPGGRHKRCFCAKVFQSDSALQIHLNSHTGERPYKCNVGNRFTT 410

Db 434 ---VTAFEAKSTSDAEAFKHKRCFCAKVFQSDSALQIHLNSHTGERPYKCNVGNRFT 489

QY 411 RGNLKVHFRHREKYPHYOMNPHVPEHLDVYITSSGLPYGMSVPERKEEAAATGGGV 470

Db 490 KGNLKVHQRHKEKIPHIOMNPPVPEHLDVNPSTGLPYGMSVSE----- 536

QY 471 ERKPLVASTTALSAATESLTLTSTAGTATAPGLPAFNKFLVMKAVEP----- 517

Db 537 -KPY--TSMIDTKFVLPVLTLTYSVGLPLPPTLPSLPFI--KTEPAPVPISSHAAPQ 589

QY 518 -KNKADEMTIP-----PGS-----EGSAT-----SGVAESSTALIMQSLKLTSL- 555

Db 590 GSVKSDSGAPDLATRNPSGVPEEVEGSAVPPFGGKGESNNAASVAPAGN--STLNSPA 648

QY 556 -----PSWALLTNHFKSTGSFPLPLCARALGASPESTSLQOLVEKIDROGAVA 604

Db 649 DGGPGGTFITNPLLLMSQFKA-----KFPFGGLDASAQASSETSLQOLVENIDKR----- 700

QY 605 VTSAAAGAPPTTSAPAPSSASSAGSNOCVICLRVLSCPRALRLHYGHGGERPKCKVGR 664

Db 701 -----ATDPMNECIIHRYVLSQSAALKMHYRTHGERPFCKICGR 740

QY 665 AFSNRGNLRHVRGKASPAARAQNSCPTQCKPTNAVTLQOHVYMHGQIPLNGSTALP 724

Db 741 AFTTKGNLKTTHSVHRAMPRLRVQHSCTPCKKFTNAVVLQOHIRMHGQIPLN--TPVP 798

QY 725 EGGGAQENGSEOSTVSGASFPOQSOQSPSPEELSEEEDEDEDEEDVDDEDLADR 784

Db 799 D-----NYPESMESDQGSF-----DEKNFDDLDNFDENME-----CE 833

QY 785 GSESGEKAIVRGDSEASGAEEVGYVAAATAGKEMDSNEKTQOSSLPPPPDLSL 844

Db 834 GSIPDTRKSADASDLSLSSPLLEMSSTIALENQMKNINAGLAEQDASL----- 884

QY 845 DQPPMEGSS-----SGVLGG--KREGKPKERSSPAS--ALPREGATSVTIVE 890

Db 885 ---KSVENGSEGDVLTNDSSVGGMESOSAGSPALISESTSSMQALSPSNS-----TQ 935

QY 891 ELSLQEAARKK-----PGSSSS----- 907

Db 936 EPIKSPGMEERQVRQGPBFGANGLSPTPVNGCALDLTSSHAETLIKEDSLGLTFPFRDQ 995

QY 908 ---RKACEVCGQAPRSQALAEHOKTHPKEGPLPTCYFCROGTELRATLKKHMLIAHVOY 964

Db 996 KFNKTACDICKTACQASALDHYRSHTERP-FICVCHNKFSTGKMLKQHMILT--HQM 1052

QY 965 -----QPPAPGQONIAALSLVPGCS--PSITST-----GISPPRKDDPT 1003

Db 1053 RDLPSQLFEPSSNLGPNQNSA--VIPANLSLTKTEVNGFVHVSPDQSDAPT 1104

Query Match	Best Local Similarity	Matches 308; Conservative 129; Mismatches 395;	Score 777.5; DB 1; Length 1355; Pred. No. 3.3e-27; Indels 471; Gaps 45;
68	ENNNSSASSEPREGHNHPVMDREHSN-PPDGGSSVPTPTPTWGPERRGSSGHFL	124	
106	ENDIKSEAKSEIEFVEDNNNNRVNMTKPSSESEKPPNAGSGMPSSPV-AEASAE	157	
125	VAAATGAAGGGGGLIILASPKIGATP-LPEESTPADPPPPPPPPPGVG	171	

[illegible]


```

Db 1184 EOLRRFAAEQOKRAAAAAAAAAAAAAAAAAQOTPPQADQDQREGGGACPPPLIMCARPPE 1243
OY 870 SSSPAAALTEPGEATSVTLVEELSLQEMARKED-----GESSKACCEVCGQAFPSQA 922
Db 1244 GMPFNLPLFPFATTQNNMCNMAMQIAQSVPAAPFNFALSGVGRS-TTCGICYKTFPCHS 1302
OY 923 ALEEHOKTHKEGFLTCVCFRCGFLERATLKKHMLAHQVO 965
Db 1303 ALEIHRSHRKERP-FKCSICDRGFTTGNLKOHLMT--HKIR 1342

RESULT 8
SALM_DROVI STANDARD; PRT; 1402 AA.
AC P39806;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Homeotic protein spalt-major.
GN SALM.
CS Drosophila villis (Fruit fly).
CC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
CC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
CC Muscomorpha; Phnydroidea; Drosophilidae; Drosophila.
CX NCBI_TaxId=7244;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94139659; PubMed=7905822;
VA Kuehnlein R.P., Frommer G., Friedrich M., Gonzalez-Gaitan M.,
VA Weber A., Wagner-Bernholz J.F., Gehring W.J., Jaechle H., Schuh R.;
VA "Spalt encodes an evolutionarily conserved zinc finger protein of
VA novel structure which provides homeotic gene function in the head and
VA tail region of the Drosophila embryo.";
VA EMOB J. 13:168-179(1994).
YC -I- FUNCTION: REQUIRED FOR THE ESTABLISHMENT OF THE POSTERIOR-MOST
YC HEAD AND THE ANTERIOR-MOST TAIL SEGMENTS OF THE EMBRYO. PROBABLY
YC FUNCTION AS A TRANSCRIPTIONAL REGULATOR. COULD REPRESS THE
YC TRANSCRIPTION OF THE TSH GENE.
YC -I- SUBCELLULAR LOCATION: Nuclear (Probable).
YC -I- DEVELOPMENTAL STAGE: FIRST EXPRESSED AT BLASTODERM STAGE AND
YC LATER IN RESTRICTED AREAS OF THE EMBRYONIC NERVOUS SYSTEM AS
YC WELL AS IN THE DEVELOPING TRACHEA.
YC -I- SIMILARITY: BELONGS TO THE SAL FAMILY OF C2H2-TYPE ZINC-FINGER
YC PROTEINS.
YC -----
YC This SWISS-PROT entry is copyright. It is produced through a collaboration
YC between the Swiss Institute of Bioinformatics and the EMBL outstation -
YC the European Bioinformatics Institute. There are no restrictions on its
YC use by non-profit institutions as long as its content is in no way
YC modified and this statement is not removed. Usage by and for commercial
YC entities requires a license agreement (See http://www.isb-sib.ch/announce/
YC or send an email to license@isb-sib.ch).
YC -----
YC EMBL; Z27444; CAA81800.1; -.
YC PIR; S42748; S42748.
YC HSSP; P15822; 1BBO.
YC FLYBase; FBgn0013137; Dvir\salm.
YC InterPro; IPR0008822; znf.C2H2.
YC Pfam; PF00096; zf-C2H2; 7.
YC PRINTS; PR00048; ZINC-FINGER.
YC SMART; SM00355; znf_C2H2; 7.
YC PROSITE; PS00028; ZINC_FINGER_C2H2_1; 7.
YC PROSITE; PS50157; ZINC_FINGER_C2H2_2; 7.
YC W Developmental protein; Zinc-finger; Metal-binding; DNA-binding;
YC Nuclear protein; Transcription regulation; Repeat.
YC T ZN_FING 432 454 C2H2-TYPE.
YC T ZN_FING 460 482 C2H2-TYPE.
YC T ZN_FING 845 867 C2H2-TYPE.
YC T ZN_FING 873 895 C2H2-TYPE.
YC T ZN_FING 905 927 C2H2-TYPE.
YC T ZN_FING 1336 1358 C2H2-TYPE.
YC T ZN_FING 1364 1386 C2H2-TYPE.
YC Q SEQUENCE 1402 AA; 153745 MW; 4CDB2EE3844F70AE5 CRC64;

```

Query Match	14.1%	Score 745.5	DB 1	Length 1402
best Local Similarity	22.1%	Pred. No. 8.5e-26		
Matches 322; Conservative 134; Mismatches 383; Indels 619; Gaps 52;				
QY 2 AHSEERSRLGVPRGEPRLGSDASEEDH-----POYCAKCOAQPTPTERFLAHQNC 54				
Db 57 ANASASAGNSCFKSPKSPQSHSEDEREPVYSQDELVEVSAOSESEIIE-----NNAD 112				
QY 55 STDPVYVVIIGGQENPNSSASSEPRGHNPNQVMDENHNP-----DSCSSVP 105				
Db 113 ETN-----ADHNHNHN-----NNKLM-----TKPVHEVEYQANLNWASMP 150				
QY 106 TDPVWGPERRKEESSGHNLYATGTAAGCGGGLIASPGLATPLREESTRAPRRPPRR 165				
Db 151 NSTT-----PPATNAVIGA-----RAQFGCATPVTLLEIQNWQMAIAQF 190				
QY 166 PPPVGVSG-----HLNPLILEEFLVLOOROIHOMQEOICROVLLG----- 209				
Db 191 AAKTIANGASGTDEAMKQIATLQTLPLNIQOQOLFQIQLIQLQSQIALNOVKONDE 250				
QY 210 -----SLQGTVGAPASBE-----LPGTASTPTPLP 238				
Db 251 ADEPLEPEREDGETPYEEBERTIADMELRQKAEKMAESKARQIHNNGVYANP--DP 309				
QY 239 LFSP-----IKPVQTKTLASSSSSSSSSSG--AEPTKCAFPHLYAPLQSGHPESA-- 287				
Db 310 SHQPNHRCRLRLKRRKEEDASAKSGASAKIFGEQESSODALNKLKEMNNPLFGPADL 369				
QY 288 -GGVGRSHKPTAPAS--PALPGSTDLASPHLAPSTTGILLAAQCLGARGLMATASPG 344				
Db 370 SSSITTHDLPENSLDLQKTOEVLDS-----ASGG 403				
QY 345 LKPKNGSGELSTGEVMP--LEKPGRAKCRFCAYKFGSDALQIHLRSHTERPYKC 401				
Db 404 IL-ANNAADDFAFGDKSSDGKGRNEPFEKHCRCYCGKVFEGSDALQIHRSHTGERPFC 462				
QY 402 NVGNSRRTETGNLKVNHNHREKYPHYQMNHPRENHDY-----I 443				
Db 463 NVGSKRPTTGNLKVNHQRNARKPRHYMNTPRLREHNDKHRRLLDOMSPDSSPTQSPA 522				
QY 444 TSSGLPYGMSVPEKA-----EEEAATPG-GGVERRPLVASTATLATESLTL 491				
Db 523 PATGLP-----PSTSTLTQQMSPMSFASAPFGLPGLIYPRM-----ELKSL 567				
QY 492 STSAGTATA-----PGLPAFNKV-----LMAVERKNADE-----NTP 526				
Db 568 GATAGTAGLPHPEFFPQMPGLGALKHTHDQSDMPGLDKSSGSPSHEEDNIATARLP 627				
QY 527 PGSE-----GSAIGVASSTATL-----MQ-----LSKIM 552				
Db 628 VNSLMEEEKETHTMEATRESAMERPLLEVRIKEERIDEDOMHLOMOKREPLTAYA 687				
QY 553 TSLPSWALL-TNH--FKSTSGFPLPLAR-ALGASPSSTK-----LOOL----- 593				
Db 688 TRHPQOCLPTTHAAKSPRSRLPLOGCARLSIMCSHPSTSMHACAVLIGSQTHTDQLTPR 747				
QY 594 -----VEKIDRQAVAVTSAASGA-----PTTSA 617				
Db 748 DNVPTMPQREDFEAFERFLPNTFTSKTDHSPIRSPAGHANAHIRSPFPNLIKHEMAFV 807				
QY 618 PAVSSAS-----SGNOCVIGLARVLSCPRALRHYQH 651				
Db 808 PRPHSDNSENENFTVSNMSETMKLELMKNKISIDNOCVQSVRYLSCSALOMNYETH 867				
QY 652 GSERPKCVGCGAFSTGRNLIKRAHFVGHKASPARAONSCPVICQKFTPAATLQONVHM 711				
Db 868 TGERPKRCICGRAFTTKGLKTHMAVHKLRPRYRNHQRVCVKKXSNALVLDQHTLH 927				
QY 712 LG-----GQIPNGSTALPEG-----GG 728				
Db 928 TGEPTDLTPEOIOAAEIRDRPRPSMBRGHFNPNPFAAFAHFNHGMAGGAGGPRGATGMGG 987				

OY 729 ----- 728
Db 988 PHNGLGSESSGGLDDNMDCCGGDDFDIDISSEHLNSNDPAATSDRRSSDDFKSLFEQ 1047
OY 729 -----AAOENGSEOSTVSGAGSPFOQOQOQ-----SPEE 758
Db 1048 KLRIPTGVVNINSHORPISAASNPNSIGSASAPASPTSSQPKPCSPYRSSCSPIR 1107
OY 759 ELSEEEDEDEDEEDVDE---DSLGRSGESGGEKAISVKG-----798
Db 1108 SVS-----ETSGALDLTFRALPPPLASSSSRSRYPOLLVSRRRRLARSVSSHRCVPMV 1162
OY 799 -----798
Db 1163 RALLSSQLPSPVIGIDCLPRLQHHLQOQHLMQOQXVAAAAAQNHHNQMHAAALH 1222
OY 799 -----DSEASGAESEVGTVAATAATGKEMDSNEKTQOSSLPPPPPSLD 845
Db 1223 QHQEHLRRAQEVQKAAQEVQKAAAAAQAQRES-----PQPPPRS-- 1269
OY 846 QPQPMQSSGVLGKKEEGKPERSSPSASALTPBG-----EATSVTLVEELS--L 894
Db 1270 -----GESSVGPAPQPNPLISARPPGMPNLPFPAPTQNMCMANQIA 1315
OY 895 QEAMRKEP-----GESSRKACEVCGQAFPSOALBEHOKTHPKEGPLTVCFCROGF 947
Db 1316 QSVMPAAPNPPLALSGVRGS--TTCGICIKTPCHSALEIHRSHTKERP-FKCNICDRGF 1373
OY 948 LERATLKKHMLAHQVO 965
Db 1374 TTKGNLKHMLT--HKIR 1389
RESULT 9
ID 2236_HUMAN STANDARD; PRT; 1845 AA.
AC O9UL37:
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Zinc finger protein 236.
GN ZNF236.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Kidney;
RX MEDLINE=99389731; PubMed=10458916;
RA Holmes D.I., Wahab N.A., Mason R.M.:
RT Cloning and characterization of ZNF236, a glucose-regulated kruppel-
RL like zinc-finger gene mapping to human chromosome 18q22-q23.";
Genomics 60:105-109(1999).
CC -1- FUNCTION: MAY FUNCTION AS A TRANSCRIPTION FACTOR.
CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A AND B (SHOWN HERE); ARE
CC PRODUCED BY ALTERNATIVE SPLICING.
CC -1- TISSUE SPECIFICITY: UBQUITOUS. EXPRESSION LEVELS ARE HIGHEST IN
CC SKELETAL MUSCLE AND BRAIN, INTERMEDIATE IN HEART, PANCREAS, AND
CC PLACENTA, AND LOWEST IN KIDNEY, LIVER, AND LUNG.
CC -1- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-
CC FINGER PROTEINS.
CC -----
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AF085244; AAD55329.1; -

DR EMBL: AF085243; AAD55328.1; -
DR HSSP: P07248; IARD.
DR Genew, HGNC:13028; ZNF236.
DR MIM: 604760; -
DR InterPro: IPR000822; Znf_C2H2.
DR Pfam: PF00096; zf-C2H2; 30.
DR Prodom: PD000003; Znf_C2H2; 5.
DR SMART: SM00355; ZNF_C2H2; 30.
DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; 30.
DR PROSITE: PS50157; ZINC_FINGER_C2H2_2; 30.
KW Transcription regulation; DNA-binding; Zinc-finger; Metal-binding;
KW Nuclear protein; Repeat; Alternative splicing.
FT DOMAIN 37 1801
FT ZN_FING 37 59
FT ZN_FING 66 88
FT ZN_FING 93 115
FT ZN_FING 121 143
FT ZN_FING 153 175
FT ZN_FING 197 219
FT ZN_FING 225 247
FT ZN_FING 253 276
FT ZN_FING 285 308
FT ZN_FING 482 504
FT ZN_FING 510 532
FT ZN_FING 538 560
FT ZN_FING 566 588
FT ZN_FING 657 679
FT ZN_FING 685 707
FT ZN_FING 713 735
FT ZN_FING 741 763
FT ZN_FING 967 989
FT ZN_FING 995 1017
FT ZN_FING 1023 1045
FT ZN_FING 1051 1073
FT ZN_FING 1166 1189
FT ZN_FING 1195 1217
FT ZN_FING 1223 1245
FT ZN_FING 1251 1273
FT ZN_FING 1657 1680
FT ZN_FING 1686 1708
FT ZN_FING 1722 1744
FT ZN_FING 1750 1772
FT ZN_FING 1778 1801
FT VARSPPLIC 1530 1558
FT VARSPPLIC 1559 1845
SQ SEQUENCE 1845 AA; 203659 MW; 2879EA91D0C6D3D8 CRC64;
MISSING (IN ISOFORM A).
VGPOCGSVEALYLENLSRKT (IN ISOFORM A).
ELNLTSGSLPSTPTSPSAISITONLVMS -> GSRSVQHS
Query Match 7.1%; Score 376; DB 1; Length 1845;
Best Local Similarity 19.2%; Pred. No. 1,5e-09;
Matches 173; Conservative 99; Mismatches 302; Indels 328; Gaps 29;
OY 364 LEKPGHKKCRFCACVFGSDSALQILHLSHTGERRPKYKCNVCGRFTTRGNLKVHFRHRE 423
Db 190 IDRGFTYSCPCGCKTFQKPSQLTGRHRIHTGERPKSCGKAFAFNKGALQTHMIKHTG 249
OY 424 KYRNV-----QMNPRVPRHL 439
Db 250 EKPRACAFCPAAFSOKGNLQSHVORVHSEVKNGPTYNCTECSCEVKSLGSLNTHISKMMH 309
OY 440 DYVITSGLPYGMVPRPEKAEEDATPGGVERKPLVASTTALSTESLTLLS-----492
Db 310 -----GFPONSTSTSTTAHVLTAFLPQLQTEAQTASASQPSQAVSDVIQQL 362
OY 493 -----TSAGTAPRG-----LPANKFVLMKAVE-----PKKADENT-PP 527
Db 363 ELSEPAVESQSPQGOQLSTVGINODILQALNSGLSIPAAHPNDSCAKTSAP 422
OY 528 GSEBSAISGVAESST-----ATLMOLSKLMTSLPS-----WA 559
Db 423 HAQNPDVSSVSEQTPDADQAEKQESPEKLDKKEKKMKKKSPFLPGSIRENGVKNH 482


```

OY 560 L-----LTNFK---STGSPPLPCARLGAASPSETS-----KLOQ 592
DB 483 VCPYCAKEFRKPSDVLVRIIRITHHEKPFKPCQCFRAFVAKSTLTAHKTHGTAKFCQY 542
OY 593 LVKIDROGAVAV-----TSAS-----610
DB 543 CMKSFSTSGSLKVIIRLHTGVRFAPCPHCKKFRSGHKKTHIASHEKHTLFRKHQOK 602
OY 611 ----GAPTSAPAPS-----SSASSGPNOCVIC 634
DB 603 PAKVRKCTKTVNPVDPDLPLOEPILITDLGLIOPIPKVOFOSYNNNNFVNADRYKCFYC 662
OY 635 LRVLSCEPRALRLHYGHGGERPFKCVKCGRAESTRGNLRAHFVGHKASPARA-----687
DB 663 HRAVKKRSCHLKOHIRSHTEKPFKSCGGRFVSAGVLKAHIFTHGLKSFKCLICNGAF 722
OY 688 -----QNSCPICQKRTFNNAVTLQOHVNMHLGGQLPNGSTALPEGGGA 730
DB 723 TTGGSLLRRHMGJHNDLRPYMCPYOCKTKTSLNCKKHKMKT-----RYELA 768
OY 731 QENGSDQSTVSGAGSPFOQSQSPSEELSEEEDEEEDVDYDDESLAGRSGSGG 790
DB 769 QOLQOHQQAASIDSTVDQSQMSTQMOVETESDELPTAEVVAANPEAMLDLEPQ---825
OY 791 EKRAISVRGDEEASGAE-----EEVGVAAAAATAGKEMDS-NEKTTQOSSLPPPPP 841
DB 826 ---HVVGTTEEAGIGQQLADPLREADEDEGFVAPODPLRGHVHVOFEEQSPAQSFEPAGLP 881
OY 842 DSL-----DQPO--PMEOGSSGVILGKKEGCKPERSSSPASALT-----PEG 881
DB 882 QGFVTYDTYHQQOFPPVQO-----LQDSSLTLESQALSTSFHQSLQAPSS 928
OY 882 EATSVT--LVELLSLOF-----AMRKEPSESSSKA--CEVCGAFPSQALEEHOKT 930
DB 929 DGMNVTYTRLIQESSQOEBLDLQAGSQFLEDNEDQRRSYRCYCKNGFKKSHLKQVRS 988
OY 931 HKKEGFLPCVPCROGFLERATLTKKHMILAHQVQFPAPHGQNTAALSLVPECSPSITS 990
DB 988 HNGEKP-YKCKLCGRFVSSGVLKSHE-KTHGTVAFS-----CSVCNMSFTT 1034
OY 991 TG 992
DB 1035 NG 1036

RESULT 10
ID ZN42_HUMAN STANDARD; PRT; 734 AA.
AC P28698; O9UBW2;
DT 01-DEC-1992 (Rel. 24, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Zinc finger protein 42 (Myeloid zinc finger 1) (MZF-1).
EN ZNF42 OR MZF1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxId=9606;
KN [1]
KP SEQUENCE FROM N.A. (ISOFORM MZFL1A).
LA MEDLINE-91317761; PubMed-1860835;
LA Hromas R., Collins S.J., Hickstein D., Raskind W., Deaven L.L.,
LA O'Hara P., Hagen P.S., Kaushansky K.;
LA "A retinoic acid-responsive human zinc finger gene, MZF-1,
LA preferentially expressed in myeloid cells.";
LA J. Biol. Chem. 266:14183-14187(1991).
LT [2]
LN SEQUENCE FROM N.A. (ISOFORM MZFLB/C).
LP TISSUE-Bone marrow;
LC MEDLINE-20432092; PubMed-10974541;
LA Peterson M.J., Morris J.F.;
LA "Human myeloid zinc finger gene MZF produces multiple transcripts and
LA encodes a SCAN box protein.";

```

```

RL Gene 254:105-118(2000).
CC -1- FUNCTION: MAY BE ONE REGULATOR OF TRANSCRIPTIONAL EVENTS DURING
CC HEMOPOIETIC DEVELOPMENT.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; MZF1A AND MZF1B/C (SHOWN HERE);
CC ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- TISSUE SPECIFICITY: PREFERENTIALLY EXPRESSED IN DIFFERENTIATING
CC MYELOID CELLS.
CC -1- INDUCTION: BY RETINOIC ACID.
CC -1- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-
CC FINGER PROTEINS.
CC -1- SIMILARITY: CONTAINS 1 SCAN BOX.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb.ch).
CC
DR EMBL; M58297; AA55898.1; -.
DR EMBL; AF055078; AAD55810.1; -.
DR EMBL; AF055077; AAD55809.1; -.
DR EMBL; AF161886; AAF80465.1; -.
DR PIR; A40751; A40751.
DR HSSP; P08047; 1SP2.
DR TRANSFAC; T00529; -.
DR GeneW; HGNC:13108; ZNF42.
DR MIM; 194550; -.
DR InterPro; IPR003309; Treg-SCAN.
DR InterPro; IPR000822; ZnF_C2H2.
DR Pfam; PF000096; zf-C2H2; 13.
DR Pfam; PF02023; SCAN; 1.
DR PRINTS; PR00048; ZNCFINGER.
DR PRODOM; PD000003; ZnF_C2H2; 7.
DR SMART; SM00431; LER; 1.
DR SMART; SM00355; ZnF_C2H2; 13.
DR PROSITE; PS50804; SCAN_BOX; 1.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 13.
DR PROSITE; PS50157; ZINC_FINGER_C2H2_2; 13.
KW Transcription regulation; DNA-binding; Zinc-finger; Metal-binding;
KW Nuclear protein; Repeat; Alternative splicing.
FT DOMAIN 44 125 SCAN BOX.
FT DOMAIN 310 321 ASP/GLU-RICH (ACIDIC).
FT DOMAIN 356 731 ZINC FINGERS.
FT ZN_FING 356 378 C2H2-TYPE.
FT ZN_FING 384 406 C2H2-TYPE.
FT ZN_FING 412 434 C2H2-TYPE.
FT ZN_FING 440 462 C2H2-TYPE.
FT DOMAIN 463 484 GLY/PRO-RICH.
FT ZN_FING 485 507 C2H2-TYPE.
FT ZN_FING 513 535 C2H2-TYPE.
FT ZN_FING 541 563 C2H2-TYPE.
FT ZN_FING 569 591 C2H2-TYPE.
FT ZN_FING 597 619 C2H2-TYPE.
FT ZN_FING 625 647 C2H2-TYPE.
FT ZN_FING 653 675 C2H2-TYPE.
FT ZN_FING 681 703 C2H2-TYPE.
FT ZN_FING 709 731 C2H2-TYPE.
FT VARSPPLIC 1 249 MISSING (IN ISOFORM MZFLB/C).
FT VARSPPLIC 250 257 EAGGTFSP -> MNGPLVYA (IN ISOFORM
FT MZFLB/C).
FT CONFLICT 304 305 AL -> RV (IN REF. 1).
SQ SEQUENCE 734 AA; 82036 MW; 2BE7D69B18F29437 CRC64;

Query Match 6.1%; Score 321; DB 1; Length 734;
Best Local Similarity 20.7%; Pred. No. 1.3e-07;
Matches 196; Conservative 74; Mismatches 309; Indels 368; Gaps 37;

OY 80 REEGHNNPOVMDT-----EHSNPPDSSGVPTPTWPERGERGESSGHFLVAATCTANG 133
DB 74 REVASKEQMLELTVLEQFLGALPPEIQARVQ-----GQRPGSPEEAALVDGLRRPBG 127

```


134 GG-----GGLIASPKLGAT---PLPEESTPAPPPPPPPPPGVSGHNLPLILE 182
128 GPRRWVTVVOVGQEVLESEKMEPSSFOPLPETEPTPEPGPTPTP----- 171
183 ELRVLOQKQI-----HOMQTEOICQVLLGSLGVGVGAPASPELPGTGAASSTKPL 237
172 --RTWQESPLGIQVNEEVEVTE-----SDFLESGLATQESV 208
238 PLFSPKIPVQTSKTLAASSSSSSSGAETPKQAFPHLYHPLGSHQPSAGVGRSHKPT 297
209 PTLPEEAORCGTVL--DQIFPHSKTGPESPSWR-----EHRALNHE--EAGCIF----- 255
298 PAPSPAL-----PGSTDQLASPHLAPSTTGL--LAQCCGAAGLQENTASPGLLK 347
256 -SPGVALGSLGSIAGSGS-----VSPHLHVPMDLGMGLSGOIQSPSR--EGGFALLL 307
348 PKNGSGELSYGEV-----MGPL-----EKPGGRH-----KCRFCAKVFGS 382
308 PSDLASEODPTDEDPGRCVGPALITTTWRSRGRHSRGRSTGGVYRGRCOVCGKVFISQ 367
383 DSALQILHLSHSGERYKCNVNGNRFTTGNLKVHFRHREKYPHYQMPHPVPHLDVY 442
368 RSNLIRHQKIHTEGRRPFCSEGRSFSRSHILRHQLTHTTEBRP-----FV 413
443 ITSSGLPYGMSVPRPEKAEEAATPGGVERKPLVASTALSTESLTLLSTAGTAPG 502
414 CGDCCGGF---VRSASLREHRRVHTG---EOPFRCAEGCGQSRQNSNLLQHORIHDDPG 467
503 LPAFNKFVLMKAVERPKKADENTPRGSESAISGVAESSTATIMOLSKIMTSLPSWALLT 562
468 -----FGAKPRAPGAPR----- 480
563 NHFKSTGSPPLPLCARALGASPSETSKLOLVEKIDROGANVNTSAASCAPTTSAPAPSS 622
481 ---PPGPFPSCSECR-----ESFARRAVLLLEHQVH-----TG 509
623 SASSGNGCVICLRVLSCPRALRYHGOHGGERPKCKVCVGFSTRGNLRAHFVGKAS 682
510 DKSFG---CVEGCEGRGRSVYLLQHRVHSGERRPACACGGSFRRNSLTQRRHRTGGE 566
683 -PAAQAQNSCPICQKKTNAVTLQOHVHMLGGQIPNGSTALPEGGAQENSGEOSTVS 741
567 RPPA-----CAEGKAPRQRTPLQHLRVHTGK-----PFACPEGC----- 603
742 GAGSFPQOQSQSPSEELSEEEDEEEDVDDESLAGSGSGKATSVRGDSE 801
604 -----QRESQRLKLTNRHQRHT-----GKK----- 623
802 EASGAEVEVGTVAATAAGKEMDSNEKTQSSLPPEPPPSLDQOPMEOGSGVILGK 861
624 -----PYNGEGGL----- 632
862 EEGKBERSSSPASALTPEGEATSVTLVEELSLQEAARKREPGSSSRKACEVCGQAPPSQ 921
633 -----GFTQVSRLT--EHQRIHTGRRPF--ACPECGQSFRQH 665
922 AALEEHOKTPKEGRLTVCVRQGFLEKATLKKHMLLHNOVPPA 968
666 ANLTQHRRIHTGERP-YACPECGKAFRQPTLQH-LRTHRRKPPA 710

RESULT 11
Y296_HUMAN
ID 1296_HUMAN STANDARD; PRT; 1829 AA.
AC 015015;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical zinc finger protein KIAA0296.
GN KIAA0296.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=97349984; PubMed=9205841;
RA Nagase T., Ishikawa K.-I., Nakajima D., Ohira M., Seki N.,
Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. VII.
RT The complete sequences of 100 new cDNA clones from brain which can
RT code for large proteins in vitro.";
RL DNA Res. 4:141-150(1997).
CC -I- FUNCTION: MAY FUNCTION AS A TRANSCRIPTION FACTOR.
CC -I- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -I- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-
CC FINGER PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb.ch/announce/>
CC or send an email to license@isb.sib.ch).
CC -----
DR EMBL: AB002294; BAA20756.1; -;
DR InterPro: IPR000822; Znf.C2H2.
DR Pfam: PF00096; Zf-C2H2; 31.
DR PRINTS; PR00048; ZNCFINGER.
DR SMART; SM00355; Znf.C2H2; 30.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 30.
DR PROSITE; PS00157; ZINC_FINGER_C2H2_2; 29.
KW Hypothetical protein; Transcription regulation; DNA-binding;
KW Zinc-finger; Metal-binding; Nuclear protein; Repeat.
FT DOMAIN 8 1783
FT ZN_FING 8 31
FT ZN_FING 48 70
FT ZN_FING 75 97
FT ZN_FING 239 261
FT ZN_FING 266 288
FT ZN_FING 294 316
FT ZN_FING 374 396
FT ZN_FING 401 424
FT ZN_FING 465 487
FT ZN_FING 492 514
FT ZN_FING 575 597
FT ZN_FING 617 639
FT ZN_FING 644 666
FT ZN_FING 821 843
FT ZN_FING 848 870
FT ZN_FING 881 904
FT ZN_FING 958 980
FT ZN_FING 1052 1074
FT ZN_FING 1079 1101
FT ZN_FING 1203 1225
FT ZN_FING 1230 1252
FT ZN_FING 1258 1280
FT ZN_FING 1299 1321
FT ZN_FING 1326 1348
FT ZN_FING 1364 1386
FT ZN_FING 1557 1579
FT ZN_FING 1585 1607
FT ZN_FING 1677 1699
FT ZN_FING 1704 1726
FT ZN_FING 1732 1754
FT ZN_FING 1761 1783
SQ SEQUENCE 1829 AA; 200760 MW; B20C482B38684895 CRC64;
Query Match 6.0%; Score 318; DB 1; Length 1829;
Best Local Similarity 20.2%; Pred. No. 5e-07;
Matches 230; Conservative 108; Mismatches 346; Indels 452; Gaps 52;

QY 28 EDHPD--VCARCKAQPTDTEFLAHQNACTDPPVWYIIGQENPNNSASSEPRP----- 81


```
2 EDTPRLSCSDCGRHPSRLBELSRHRELHPSR-----NQDSEADSLPRRYRCQ 51
82 ---BEHNNPQVNDTEHNSPPDSSSVPTDPTWGPBERGEESGHLVATGTAAGGGGI 138
52 QCGRYRHP-----GSLV-----NHRRTHTGTPCTTCG-----81
139 ILASFKLATPLRPES---TPAPP-----PPPPPPPVGGSGHLNPLILELKVLOQ 189
82 ---KQPSNPAKLSHMTHTAPBEGRRHRRPRRKEATP-----HLO-----118
190 ROIHOMQMTQICROVLLLSLIGQITWGAAPASPSLPGTITASSSTKPLPLFSPKPVQTS 249
119 -----GEVSTSDSWCQRLGS-----SEGWENQFK-----HTE 145
250 KTLIASS-----SSSSSSSGAETPRKQAFPHLYHPLGSOHP-----FSAGVYGRSHK-----PTP 296
146 ETPDCEVYDPDPAASGTWEDLPTKOREGL-----ASHRPEDGADGWGSPSTNSARAPLP 200
299 APSPALPSTQDLIASPHLAFPTSTGILAAQCLGAAGLAEATASFGLLPKNGSGELSYG 358
201 IPASLSLNLLEGYLAESVNF-----TG-----GQETQSP-----231
359 EVMGPLEPFGCGHNRKCFKAKVNGSDSALOIHLSHGBRPPYKCNVCNRFTRGNLKVHF 418
232 ---PAEBE-RRKYCSQCKYKHAAGSLTNHROSHT-LGIYPCALCFKFEFSMIMALKNHS 285
419 HRHREKYHVQNNPHVPHLDVYITSSGLPYGKSVPEKAEAEATPGGVERKPLVAS 478
286 RL-----HAQRYRPHCPH-----CPRYRPLRELLHHOOSHGEORPPR-----324
479 TTALSTATESLTLTSSAGTAPALPAFNKFLVAKAVEPRKNKADENTPPGSGAISGVA 538
325 -----WEEKMPTTNGHT-----DESSODOLPSAQ--MLNGSA 355
539 ESSATLMLQSLMTSLP-----SMALLTNHKS--TGSEFPLICARALGASS 585
356 ELSTSGELEDSELETRPRPCGCGRTYRHAGSLINRKSQOTGYVPSLCSKOLFNAA 415
586 ETSKLOQLVEKIDROGAVAVTASAGAPTTSAAPASSSSSGPNQVYCLRYLSCPRLR 645
416 LKNHVR--AHNHRPGV-----GENGPVSV-PRAP-----LL 444
646 LHYGOHGE-----RPFKYVCGRASTGRNLAHFVGHKASPARAONSCPICOK 696
445 LAETTHKHEEDPTTLDRHRYKCEGGRAYRHRSLSLVNHRHSHT-----GEYQCSLCP 499
697 KFTNAVTLQOHNVRMLHGGIPINGGTALPEGGGAQENGSEOSTVSAGSPQOQSOQ--P 754
500 KYPNLMALRNHVRHC-----KAARRSADIGAGARSHLKVLELPP 539
755 SPEERLSEEEED-----EEBEDVT-----DEDSLGRGSESG- 789
540 DPVBAEAPRHTQDHYVCKHEEEDITTPADKTAAHICSLGLFEDAESLERHGLTHGA 599
790 GEK-----ALSVR-----GD-----799
600 GEKENSRTETWSPRAPEFCROCGKSYRHSGLINHRQTHQTFSCGACAKHFTMAAM 659
800 -----SEASGAEEVGTVAATAAGKEMDNSEKTTQOSLPPPPP 841
660 KNLHRRHSRRSRHRRKRGASGGRG--AKLLAAESWTRLEEDNE-----GLESPODP 711
842 -----DSLDPQRPWEGSGSGVLGCKEGBKPP-----RSSSPASA 876
712 SGSEPHGAEGLNLESDGDLQAESEGDCKGLERDETHFGQDKESGGTGGLERKDALSLDN 771
877 LTPRGE-----ATSVTLVEELSLQEAAMKEPGEES-----SKKACVCGQA 917
772 LDIPGEEGGTHFCDSLJGVDE--DQKPRATGPNSSSHSANAATGWAQAHTCSDDCGHS 829
918 FPSOALAEHOKTHPRKEGPLFTCVCFROGFLERATLKKHMLLAHQ-----VOPF 967
```

```
DB 830 FPHATGLLSHRPCHP--GIYQCSLCPKFPDSLPAIRSH--FQNNRPGEATSAQPF 881
RESULT 12
ID 2341_HUMAN STANDARD; PRT; 773 AA.
AC Q95YN7; Q95Y75;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Zinc finger protein 341.
GN ZNF341.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21638749; PubMed=11780052;
RA Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R.,
RA Jones M., Stavrides G., Almeida J.P., Babage A.K., Baggley C.L.,
RA Bailey J., Barlow K.F., Bates K.N., Beard L.M., Beare D.M.,
RA Beasley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J.,
RA Buck D., Burrill W.D., Butler A.P., Carder C., Carter N.P.,
RA Chapman J.C., Clamp M., Clark A.P., Clark L.N., Clark S.Y., Clee C.M.,
RA Clegg S., Cobley V.E., Collier R.E., Connor R.E., Corby N.R.,
RA Coulson A., Coville G.J., Deadman R., Dhami P.D., Dunn M.,
RA Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,
RA Grafham D.V., Griffiths C., Griffiths M.N.D., Gwilliam R., Hall R.E.,
RA Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J.,
RA Huckle E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D.,
RA Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,
RA Levasalho M.H., Leverkus M.A., Lloyd C., Lloyd D.M., Lovell J.D.,
RA Marsh V.L., Martin S.L., McConachie L.J., McIay K., McMurtry A.A.,
RA Milne S.A., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T.,
RA Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I., Ramsay H.,
RA Phillimore B.J.C.T., Pratchalingam S.R., Plumb R.W., Ramsay H.,
RA Rice C.M., Ross M.T., Scott C.E., Sehra H.K., Shownkeen R., Sims S.,
RA Skuce C.D., Smith M.L., Soderlund C., Steward C.A., Sulston J.E.,
RA Swann R.M., Symamore N., Taylor R., Tee L., Thomas D.W., Thorpe A.,
RA Tracey A., Tromans A.C., Vaudin M., Wall M., Wallis J.M.,
RA Whitehead S.L., Whitaker P., Willey D.L., Williams L., Williams S.A.,
RA Wilmink L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,
RA Rogers J.;
RT "The DNA sequence and comparative analysis of human chromosome 20.";
RL Nature 414:865-871(2001).
RN [2]
RP SEQUENCE OF 227-773 FROM N.A.
RA Isegal T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Mshikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,
RA Wagaatsuma M., Hosoliri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
RA Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,
RA Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K., Masuo Y.,
RA Ninomiya K., Iwayanagi T.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
CC -I- FUNCTION: MAY FUNCTION AS A TRANSCRIPTION FACTOR.
CC -I- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -I- SIMILARITY: BELONGS TO THE KRUPPEL FAMILY OF C2H2-TYPE ZINC-
CC FINGER PROTEINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-stb.ch/announce/
CC or send an email to license@isb-stb.ch).
CC
CC EMBL: AL050349; CAC14087.2; -
CC EMBL: AK027550; BAB55193.1; -
CC HSSP: P07248; ZADR.
```


KW	Transcription regulation; DNA-binding; Zinc-finger; Metal-binding;
KM	Nuclear protein; Repeat.
FT	DOMAIN 57 520 ZINC FINGERS.
FT	ZN_FING 57 79 C2H2-TYPE.
FT	ZN_FING 85 107 C2H2-TYPE.
FT	ZN_FING 163 186 C2H2-TYPE.
FT	ZN_FING 274 296 C2H2-TYPE.
FT	ZN_FING 302 324 C2H2-TYPE.
FT	ZN_FING 498 520 C2H2-TYPE.
FT	CONFLICT 232 233 MISSING (IN REF. 2).
FT	CONFLICT 436 436 O -> E (IN REF. 2).
ISO	SEQUENCE 722 AA; 76876 MW; B19DA77B148BC45B CRC64;
Query Match	5.7%; Score 303; DB 1; Length 722;
Best Local Similarity	22.4%; Pred. No. 9e-07;
Matches 164:	Conservative 70; Mismatches 271; Indels 228; Gaps 27;
2Y	348 PKNGSGELSYGVMGLEKGGKRN-KCRFCAKVFGSDSALOHLHSHGPERPKVCNCGN 406
Db	33 PAVSAGSLGMGAVSWSESXGGERFPCVCGKRRFNSTLALHRAHPGAQAFQCPHCGH 92
2Y	407 RFTTGNLKVHFHREKYPHYQMNPHVPEHLDYVITSSGLPYGMSVPEKAEAEATP 466
Db	93 RAAQALRLSRSLRTHQPERP----- 112
2Y	467 GGGVERKPLVASTTALSAVESLTLTSTAGTATAPGLPAFNKFLMKAVEPKKADENTP 526
Db	113 -----RSP--AARLLELEBRALLREARLGRARSSG-----GMOAT 146
2Y	527 PGSESSAISGVNESSATITMQLSKMTSLPSSWALLTNHFKSTGSPPLPCARALGASPSF 586
Db	147 PATEGLARPOADSSS-----AFRCPCYCKKERTS-AE 177
2Y	587 TSKLQOLVEKIDROGAVAVTSA-----ASGAP-----TTSAPAPSSASSGPN 629
Db	178 RERHLHLHRPKKCGICSSSOEELHLSLTAHGAPERPLAATSAPPPQPPQPPQ 237
2Y	630 QCVICLRVLSCPRALRLHYGQNGER-----PFCKVCGRASFSTRGN 671
Db	238 -----PEPRSVQPPREPEREATPTPAPAPREPPAPPEFCQVCGSQSFTQSWF 288
2Y	672 LRAHVGHKASPAARQNSQCTOKFTNAVTLQOQVNMH-----LGQITNGGATL-- 723
Db	289 LKGNHRKHKAS----FDHACPCVGRCKEKFEMFLKNHMKVHASKIGPLRLAPGAPSGPARAP 344
2Y	724 --PEGGAQOENGSEOSTVSGASFPQOOSQPSPEBELSEEEDEDEEDVTEDSL 781
Db	345 QPDLGLALAYEPLRALLLAPPT-PAERREPPSLGLYLSLRAGGRP-----NEGAE 397
2Y	782 AGRGSESG-----EKAISVAGDSEBAGAEENVTAANA---TAGKEMDSNEKTTQSS 834
Db	398 PGPGRSFGGFRPLSSALPARARRHRAEEREEVEVLEAOEFTWARGRISGS-----LAS 451
2Y	835 LPPPPPSLDQPPWEGQSSGVLGKKEGGKPKRRSSSPASALPREGATSTIVIELS- 893
Db	452 LHRP-----GEGPCHSASAQA---OAKSTATQENGL 482
2Y	894 LQEAARKPEGESSSRKACVCGAPFSQALAEHOKTHPKRGPLFTCYFCROGFLEATL 953
Db	483 LVGGRTRPBGGRGATGKDCFCGCSFSAHNLKVLRYHTGERP-YKCHNDOVAAGQSSGL 541
2Y	954 KKHMLAHQOV-----PRAP-----HGFONTAALSLVP-----GCSPTSTSG 992
Db	542 KYH-LQRHHRDRSGAGPPRPSPPSQGSAAQSAKPSQAPATWEGASSPPRSSG 600
2Y	993 LSPPRKDDPTIP 1005
Db	601 AGPGSR-KPASP 612
RESULT 14	
N84_HUMAN	
D_ZN84_HUMAN	STANDARD; PRT; 738 AA.

AC	P51523; Q9NNX7; Q9UC17; Q9UC18;
DT	01-OCT-1996 (Rel. 34, Created)
DT	01-OCT-1996 (Rel. 34, Last sequence update)
DT	15-JUN-2002 (Rel. 41, Last annotation update)
DE	Zinc finger protein 84 (Zinc finger protein HPR2).
GN	ZNFR84.
OS	Homo sapiens (Human).
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
OX	NCBI_TaxID=9606;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	TISSUE=Placenta;
RA	MEDLINE=89377476; PubMed=2505992;
RA	Bellefroid E.J., Lecocq P.J., Benhida A., Poncelet D.A.,
RT	Belayew A., Martial J.A.;
RT	"The human genome contains hundreds of genes coding for finger
RL	proteins of the Kruppel type.";
RL	DNA 8:377-387(1989).
RN	[2]
RP	SEQUENCE FROM N.A.
RA	Grimaldi G.;
RL	Submitted (Feb-2002) to the EMBL/GenBank/DBJ databases.
RN	[3]
RP	SEQUENCE OF 1-257 FROM N.A.
RC	TISSUE=Teratocarcinoma;
RA	MEDLINE=92051312; PubMed=1945843;
RA	Rosati M., Marino M., Franze A., Tramontano A., Grimaldi G.;
RT	"Members of the zinc finger protein gene family sharing a conserved N-
RT	terminal module.";
RL	Nucleic Acids Res. 19:5661-5667(1991).
CC	-1- FUNCTION: MAY BE INVOLVED IN TRANSCRIPTIONAL REGULATION.
CC	-1- SUBCELLULAR LOCATION: Nuclear (Probable).
CC	-1- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-
CC	FINGER PROTEINS.
CC	-1- SIMILARITY: CONTAINS 1 KRAB DOMAIN.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC	the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC	or send an email to license@isb-sib.ch).
CC	-----
DR	EMBL: M27878; AAA79359.1; -
DR	EMBL: X60156; CAB94232.2; -
DR	HSSP: P08047; ISP2.
DR	Genew: HGNC:13159; ZNF84.
DR	InterPro: IPR001909; KRAB.
DR	InterPro: IPR000822; Znf_C2H2.
DR	Pfam: PF00096; zf-C2H2; 19.
DR	Pfam: PF01352; KRAB; 1.
DR	PRINTS: PR00048; ZINCFINGER.
DR	ProDom: PD00003; Znf_C2H2; 14.
DR	SMART: SM00349; KRAB; 1.
DR	SMART: SM00355; Znf_C2H2; 19.
DR	PROSITE: PS50805; KRAB; 1.
DR	PROSITE: PS00028; ZINC_FINGER_C2H2_1; 19.
DR	PROSITE: PS50157; ZINC_FINGER_C2H2_2; 19.
KW	Transcription regulation; DNA-binding; Zinc-finger; Metal-binding;
KW	Nuclear protein; Repeat.
FT	DOMAIN 8 79 KRAB.
FT	DOMAIN 207 733 ZINC FINGERS.
FT	ZN_FING 207 229 C2H2-TYPE.
FT	ZN_FING 235 257 C2H2-TYPE.
FT	ZN_FING 263 285 C2H2-TYPE.
FT	ZN_FING 291 313 C2H2-TYPE.
FT	ZN_FING 319 341 C2H2-TYPE.
FT	ZN_FING 347 369 C2H2-TYPE.
FT	ZN_FING 375 397 C2H2-TYPE.
FT	ZN_FING 403 425 C2H2-TYPE.
FT	ZN_FING 431 453 C2H2-TYPE.

FT	ZN_FING	459	481	C2H2-TYPE.
FT	ZN_FING	487	509	C2H2-TYPE.
FT	ZN_FING	515	537	C2H2-TYPE.
FT	ZN_FING	543	565	C2H2-TYPE.
FT	ZN_FING	571	593	C2H2-TYPE.
FT	ZN_FING	599	621	C2H2-TYPE.
FT	ZN_FING	627	649	C2H2-TYPE.
FT	ZN_FING	655	677	C2H2-TYPE.
FT	ZN_FING	683	705	C2H2-TYPE.
FT	ZN_FING	711	733	C2H2-TYPE.
FT	CONFLICT	62	E -> Q (IN REF. 2 AND 3).	
SO	SEQUENCE	738 AA;	85458 MW;	55268AID98179EED CRC64;
Query Match				
Best Local Similarity		5.7%;	Score 301.5;	DB 1; Length 738;
Matches 133; Conservative		21.5%;	Pred. No. 1.1e-06;	
		67;	Mismatches 235;	Indels 185; Gaps 18;
OY	365	EKPGRHKRCFCAYKFGSDALDHLNSHTGERPYKVCNVCNFTTGNLKVHFRHR-E	423	
Db	260	EKP---YNSOGCKAFSSQKSQSLTSHQRTHTGKPYECGECGCKAFSRKSHLISHWRHTGE	316	
OY	424	KYPHVQMPHPVPHLDVYITSSGLPYGMSVPRKAEDEAATPGGVERKPLVASTTALS	483	
Db	317	K-----PYGCNECGRAFSERK-----	331	
OY	484	ATESLTLSTAGTATAPGLPAFNKFLMKAVEPKNKADENTPPGSEGAISGVASSTA	543	
Db	332	-----SNLTNHQRTHTGKPEECRECKAFSRKS-----	360	
OY	544	TLMOQLSKMTSLPSMALLNHFKESTGSPPLPL-CARAL-----GASPSSET	587	
Db	361	-----OLVTHHRTHTGKRPFGCSOGRKAFPEKSELIRHOTHTGKPYEC	405	
OY	588	SKLOOLVERKIDROGAVALVSAAGAPTTAPAPSSASSGPRQCVICLAVLSCPRALRLH	647	
Db	406	SECRKAFR--ERSLSLNHQRTHTG-----EKPRHCIOGCKAFSSQKSQSLTSH	449	
OY	648	YGOHGERPFKCVCGRAFTSGRLRAHFVGHKASPARAONSCPIQCKKFTNAVTLOOH	707	
Db	450	QMTHTGKPFKSCGKAFSRKSQSL---YRHQRTHTGKPYECSECGKAFSEKSLITNH	505	
OY	708	VMNHLCGQIPNGSTALPREGGAOENSGSEOSTVSGAGSPFOOQSOQSPPEELSEEE	767	
Db	506	QRH-----TGEKPYVSCGCKAFSCCKSHLISHQRTHTGKPYE	544	
OY	768	DEEEDVDVDESLAGRSSESGEKAISVRGDSERASGAEEVGYAAATATAGKENDSNE	827	
Db	545	CSECGKAFGEKSSLSLATQRTHTGKPYECR-DCEKAFSOKSOLMT-----HQRTHTGE	596	
OY	828	KTTQGSLLPPRPDSLDOPQPMGOGSSGVLGKREGGKPERSSSPASALTPEGATSVT	887	
Db	597	K-----PYCSLCRAFFKSELSIRHLRTHHTGKPYECNECKRAFR-----	637	
OY	888	LVEELSLQEAARKPEGSSSRKACVCGQAFPSQALAEHQRTHTPEKGLFTGCVCRQGF	947	
Db	638	--EKSSLINHQRTHHTGKPRE--CSECGKAFSRKSHLIRHQRTHTGKPE--YGCSECKRAF	692	
OY	948	LEBATLKKHMLLHHQVOPF	967	
Db	693	SQKSQSLVNHQRT-HTGKPY	711	
RESULT 15				
ID	XFIN_XENLA	STANDARD;	PRT;	1350 AA.
AC	P08045;			
DT	01-AUG-1988	(Rel. 08, Created)		
DT	01-AUG-1988	(Rel. 08, Last sequence update)		
DT	16-OCT-2001	(Rel. 40, Last annotation update)		
DE	zinc finger protein xfin.			
GN	XFIN.			
OS	Xenopus laevis (African clawed frog).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			

OC	Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
OC	Xenopodinae; Xenopus.
OX	NCBI_TaxID=8355;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	TISSUE=Embryo;
RX	MEDLINE=88082679; PubMed=2826129;
RA	Ruiz I Altaba A.; Perry O'Keefe H.; Melton D.A.;
RT	"Xfin: an embryonic gene encoding a multifingered protein in
RT	Xenopus";
RL	EMBO J. 6:3065-3070(1987).
RN	[2]
RP	CHARACTERIZATION.
RX	MEDLINE=94021366; PubMed=7692399;
RA	Andreazoli M.; de Lucchini S.; Costa M.; Baracchi G.;
RT	"RNA binding properties and evolutionary conservation of the Xenopus
RT	multifinger protein Xfin";
RL	Nucleic Acids Res. 21:4218-4225(1993).
RN	[3]
RP	STRUCTURE BY NMR OF FINGER 31.
RX	MEDLINE=89346749; PubMed=2503871;
RA	Lee M.S.; Gipeert G.P.; Soman K.V.; Case D.A.; Wright P.E.;
RT	"Three-dimensional solution structure of a single zinc finger DNA-
RT	binding domain";
RL	Science 245:635-637(1989).
RN	[4]
RP	STRUCTURE BY NMR OF A FINGER.
RX	MEDLINE=89378224; PubMed=2506074;
RA	Lee M.S.; Cavanaugh J.; Wright P.E.;
RT	"Complete assignment of the 1H NMR spectrum of a synthetic zinc
RT	finger from Xfin. Sequential resonance assignments and secondary
RT	structure";
RL	FEBS Lett. 254:159-164(1989).
CC	-1- FUNCTION: BINDS RNA. COULD FUNCTION IN POST-TRANSLATIONAL
CC	REGULATION PROCESSES.
CC	-1- SUBCELLULAR LOCATION: Cytoplasmic.
CC	-1- TISSUE SPECIFICITY: OOCYTES AND IN SPECIALIZED CELL TYPES SUCH
CC	AS NEURAL RETINA CONES.
CC	-1- DEVELOPMENTAL STAGE: EXPRESSED THROUGHOUT OOGENESIS AND
CC	EMBRYOGENESIS.
CC	-1- DOMAIN: CONTAINS 37 FINGER MOTIFS IN 6 DOMAINS.
CC	-1- PTM: PHOSPHORYLATED.
CC	-1- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-
CC	FINGER PROTEINS.
CC	-1- SIMILARITY: CONTAINS 1 KRAB DOMAIN.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC	the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC	or send an email to license@isb-sib.ch).
CC	-----
DR	EMBL; X06021; CAA29425.1; -
DR	PIR; S00647; S00647.
DR	PDB; 1ZNF; 15-OCT-91.
DR	InterPro; IPR001909; KRAB.
DR	InterPro; IPR000822; znf_C2H2.
DR	Pfam; PF00096; zf-C2H2; 36.
DR	Pfam; PF01352; KRAB; 1.
DR	PRINTS; PR00048; ZINC_FINGER.
DR	ProDom; PD000003; znf_C2H2; 20.
DR	SMART; SM00349; KRAB; 1.
DR	SMART; SM00355; znf_C2H2; 35.
DR	PROSITE; PS50805; KRAB; 1.
DR	PROSITE; PS00028; ZINC_FINGER_C2H2_1; 35.
DR	PROSITE; PS50157; ZINC_FINGER_C2H2_2; 37.
KW	zinc-finger; Metal-binding; RNA-binding; Repeat; 3D-structure;
KW	phosphorylation.
FT	DOMAIN 1 58 KRAB.
FT	DOMAIN 108 298 ZINC-FINGERS I.
FT	DOMAIN 326 488 ZINC-FINGERS II.


```

T DOMAIN 503 721 ZINC-FINGERS III.
T DOMAIN 750 940 ZINC-FINGERS IV.
T DOMAIN 988 1066 ZINC-FINGERS V.
T DOMAIN 1136 1298 ZINC-FINGERS VI.
T ZN_FING 108 130 C2H2-TYPE.
T ZN_FING 136 158 C2H2-TYPE.
T ZN_FING 164 186 C2H2-TYPE.
T ZN_FING 192 214 C2H2-TYPE.
T ZN_FING 220 242 C2H2-TYPE.
T ZN_FING 248 270 C2H2-TYPE.
T ZN_FING 276 298 C2H2-TYPE.
T ZN_FING 326 348 C2H2-TYPE.
T ZN_FING 354 376 C2H2-TYPE.
T ZN_FING 382 404 C2H2-TYPE.
T ZN_FING 410 432 C2H2-TYPE.
T ZN_FING 438 460 C2H2-TYPE.
T ZN_FING 466 488 C2H2-TYPE.
T ZN_FING 503 525 C2H2-TYPE.
T ZN_FING 531 553 C2H2-TYPE.
T ZN_FING 559 581 C2H2-TYPE.
T ZN_FING 587 609 C2H2-TYPE.
T ZN_FING 615 637 C2H2-TYPE.
T ZN_FING 643 665 C2H2-TYPE.
T ZN_FING 671 693 C2H2-TYPE.
T ZN_FING 699 721 C2H2-TYPE.
T ZN_FING 750 772 C2H2-TYPE.
T ZN_FING 778 800 C2H2-TYPE.
T ZN_FING 806 828 C2H2-TYPE.
T ZN_FING 834 856 C2H2-TYPE.
T ZN_FING 862 884 C2H2-TYPE.
T ZN_FING 890 912 C2H2-TYPE.
T ZN_FING 918 940 C2H2-TYPE.
T ZN_FING 988 1010 C2H2-TYPE.
T ZN_FING 1016 1038 C2H2-TYPE.
T ZN_FING 1044 1066 C2H2-TYPE.
T ZN_FING 1136 1158 C2H2-TYPE.
T ZN_FING 1164 1186 C2H2-TYPE.
T ZN_FING 1192 1214 C2H2-TYPE.
T ZN_FING 1220 1242 C2H2-TYPE.
T ZN_FING 1248 1270 C2H2-TYPE.
T ZN_FING 1276 1298 C2H2-TYPE.
T ZN_FING 1045 1045 C2H2-TYPE.
T STRAND 1052 1052
T STRAND 1056 1065
T HELIX 1066 1066
T TURN 1066 1066
Q SEQUENCE 1350 AA; 155804 MW; 27F10AB0851E0AD8 CRC64;

Query Match 5.6%; Score 296.5; DB 1; Length 1350;
Best Local Similarity 21.3%; Pred No.3.2e-06;
Matches 131; Conservative 64; Mismatches 249; Indels 171; Gaps 17.

Y 365 EKGGRHKRCRCACAVFGSDSALOHLRSHTGERPYKCNVCGRFTTNGNLKVHFRHREK 424
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Y 407 EKP---FKSCHDKKFTERSALAKHQRTHGKPYKSCDCKEFTQSRNLILHQRIHGE 463
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Y 425 YPHQMNHPVDEHDIYTTSSGLPRFGMSVPRPEKADEEATPGGGEKPLVASTTALSA 484
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Y 464 RPY----- 466
Y 485 TESTLLSTAGTATAPGIPANFKFVLMKAVEPKNKADENTPPGSEGSAGISVAESSTA- 543
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Y 467 --KCTLCORT-----FIQNSDLVVKHQKHANIP-----LSDPHTAN 500
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Y 544 TLNQLSKIMTSLPMSALLTNHFKSTGSPFLPLCARALGASPESTSKLOOLEVEKIDROGAV 603
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Y 501 SPHKSCKDLTFESHSTFWKHSKISGKKFKQCAEC-----KKGFQKSDLVKHI 550
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Y 604 AYVSAASGAPITTSANAPSSASGPNQVICYLFLVLSCPRALRLHYGHHGGRPPKCYCG 663
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Y 551 RVHNGEK-----PKCLCKKRSFSQNSDLHKHWRILHTGEKPPPCYTCDD 593
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Y 664 RAESTRGNLRAHFVGHKASPARAONSCPIQCKFTNAVTLQOHVHMHLLGQIIPNGTAL 723
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```

Db 594 KSTERSALIKH---HRTHTGER-PHKSCVCQKGFIOKSLATLKHSRTH----- 637
Qy 724 PBGGGAQAEQSGSQSYSGAGSPFOQSQSPREELSEEEEDDEEDDYTDSDSLAG 783
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 638 -----TGERPYPTQCGKSFQNSDLVKHQRHHTGKPYHCTCKNRFTLEGGSLVK 688
Qy 784 RSESGGEKALISVRCDSSEESGAAEEV-----GTVAATATAGKEMDNKTTQSSLP 836
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 689 HRTHTGKPYRCPQCKRTIQSSDLVKHLVHNGENPRAAFHEILLRREULTSRSPD 748
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 837 PPPPDSLDQPPQMEGSSGVLGKEGEGKPERSSPASALT----PEGEATSVTLVEEL 892
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 749 PYPCTEC-----GKVFHQRPALILKHLRHTKTRKPYPCNECDK-SFFQTS 791
Qy 893 SIQEARKEKPESSSRKACEVCGAFAFPOALAEHOKTHPKGPIPTCYFCROGFLERAT 952
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 792 DLVKHLRHTGKPYR--CEPCNKGFQNSDLVKHQRHHTGERP-YTCSQCKGFLORSA 848
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 953 LKKHMLLAHHQVQPF 967
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 849 LTKHM-RTHDGEKPY 862

RESULT 16
ZN91_HUMAN STANDARD; PRT; 1191 AA.
ID ZN91_HUMAN
AC 005481.
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Zinc finger protein 91 (Zinc finger protein HNF10) (HPF7).
GN ZNF91.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RX MEDLINE FROM N.A.
RX MEDLINE=93223677; PubMed=8467795;
RA Bellefroid E.J., Marine J.C., Ried T., Lecocq P.J., Riviere M.,
RA Amemiya C.T., Poncelet D.A., Coulle P.G., de Jong P.J.,
RA Szpieter C., Ward D.C., Martial J.A.;
RT "Clustered organization of homologous KRAB zinc-finger genes with
RT enhanced expression in human T lymphoid cells.";
RL EMBO J. 12:1363-1374(1993).
RN [2]
RP SEQUENCE OF 15-204 FROM N.A.
RX MEDLINE=91219421; PubMed=2023909;
RA Bellefroid E.J., Poncelet D.A., Lecocq P.J., Revelant O.,
RA Martial J.A.;
RT "The evolutionarily conserved Kruppel-associated box domain defines a
RT subfamily of eukaryotic multifingered proteins.";
RL Proc. Natl. Acad. Sci. U.S.A. 88:3608-3612(1991).
CC -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -!- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-
CC FINGER PROTEINS.
CC -!- SIMILARITY: CONTAINS 1 KRAB DOMAIN.
CC -!- CAUTION: THE SEQUENCE FROM POSITION 1159 TO THE C-TERMINAL IS
CC DERIVED FROM THE TRANSLATION OF AN ALT REPEAT.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; L11672; AA59469.1; -
DR EMBL; M61871; AA58672.1; ALT_SEQ.
DR PIR; F39384; F39384.
DR PIR; S35305; S35305.
DR HSP; P08047; 1SP2.
```


[illegible]

```

Db 757 LTKHKRIHREKPKCKECCGKAFIMWSLTITNKKRIHTEGKPKCECGCAFAFSSSTLTGH 816
QY 589 KLQQLVERKIDRCQAVAVTSAASGAPTTSAAPSSSSASSGPNOCVTCRLVLSCPRALRLHY 648
Db 817 KTIHTGEK-----PYCKCECGKAFKHSALAKHK 845
QY 649 GONGGERPKCKKVCAGFAFSTRGNIRAFVGHKAKSPARAQNSCPIQCKKFTNMTVTLQOHV 708
Db 846 ITHAEKLYKCECGKAFNOSNLTITKRIHTKPKSKSEE---CDKAFIMWSSTLTGKH 901
QY 709 RMHLGQIPINGCGTALPEGGAQENGSEOSTVSGASFPQOQSOPSPPEELSEEEED 768
Db 902 RHH-----TREPKPKCECGCAFS-----QPSHLTHHKRMHTEGKPKYKC 940
QY 769 EEEEDVDTEDSLARGSESGEKEAISVRGDSSEASGAEEVGYVAAATAGKEMDSNE- 827
Db 941 EECGKAFSOSTLTITKRIHTEGKPKYK---CECGKAFKRSSTLT---TEHKRIHTEGK 992
QY 828 -----KTOQSSILPPIPPPPDLDQOPMEQSSGVILGKKEGGKPERSSSPASALTLP 879
Db 993 PKYCECGKAFSOS-----TLTRITRNHTGE-----KPKYCECGKAFNR 1033
QY 880 EGEATVTIVEEELSLQAMRKPEGSSSRKACEVCGAPPSOALEENOKTHPKEGPLFT 939
Db 1034 SKRLTHKII-----HTEGKPKYK---CECGKAFISSSTLTGNKKRIHREKPK-YK 1079
QY 940 CVFCROGFLEKATLKKHMLLAHNOVQPF 967
Db 1080 CEECGKAFSOSTLTITRHKRL-HTGEKPKY 1106

RESULT 17
HBL1_CAEEL
AC 09XYD3: Q19389; STANDARD; PRT; 982 AA.
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hunchback-like protein.
GN HBL-1 OR F13D11.2.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=99117349; PubMed=9917360;
RA Fay D.S., Stanley H.M., Han M., Wood W.B.;
RT "A Caenorhabditis elegans homologue of hunchback is required for late
RN stages of development but not early embryonic patterning.";
RL Dev. Biol. 205:240-253(1999).
RP [2]
RC SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Fulton L.;
RL Submitted (SEP-1999) to the EMBL/Genbank/DBJ databases.
CC -! FUNCTION: REQUIRED FOR LATE STAGES OF DEVELOPMENT.
CC -! SUBCELLULAR LOCATION: Nuclear (Probable).
CC -! TISSUE SPECIFICITY: EXPRESSED PRIMARILY IN ECTODERMAL CELLS DURING
CC EMBRYONIC AND LARVAL DEVELOPMENT.
CC -! SIMILARITY: BELONGS TO THE HUNCHBACK FAMILY OF C2H2-TYPE ZINC-
CC FINGER PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation.
CC The European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/announcements
CC or send an email to license@ebi.ac.uk).
CC -----
DR EMBL; AF097737; AAD16170.1; -

```



```
DR EMBL: AF277623; AAK61306.1; -
DR EMBL: AB046831; BAB13437.1; -
DR EMBL: AK024442; BAB15732.1; -
DR HSSP: P08047; 1SP2
DR InterPro: IPR001909; KRAB
DR InterPro: IPR000822; znf_C2H2.
DR Pfam: PF00096; zf-C2H2; 20.
DR Pfam: PF01352; KRAB; 1.
DR PRINTS: PR00048; ZINC_FINGER.
DR PRODOM: PD000003; znf_C2H2; 2.
DR SMART: SM00349; KRAB; 1.
DR SMART: SM00355; znf_C2H2; 20.
DR PROSITE: PS00805; KRAB; 1.
DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; 20.
DR PROSITE: PS00157; ZINC_FINGER_C2H2_2; 20.
KW Hypothetical protein; transcription regulation; DNA-binding;
KW Zinc-finger; Metal-binding; Nuclear protein; Repeat.
FT DOMAIN 8 79 KRAB.
FT ZN_FING 257 811 ZINC_FINGERS.
FT ZN_FING 257 279 C2H2-TYPE.
FT ZN_FING 285 307 C2H2-TYPE.
FT ZN_FING 313 335 C2H2-TYPE.
FT ZN_FING 341 363 C2H2-TYPE.
FT ZN_FING 369 391 C2H2-TYPE.
FT ZN_FING 397 419 C2H2-TYPE.
FT ZN_FING 425 447 C2H2-TYPE.
FT ZN_FING 453 475 C2H2-TYPE.
FT ZN_FING 481 503 C2H2-TYPE.
FT ZN_FING 509 531 C2H2-TYPE.
FT ZN_FING 537 559 C2H2-TYPE.
FT ZN_FING 565 587 C2H2-TYPE.
FT ZN_FING 593 615 C2H2-TYPE.
FT ZN_FING 621 643 C2H2-TYPE.
FT ZN_FING 649 671 C2H2-TYPE.
FT ZN_FING 677 699 C2H2-TYPE.
FT ZN_FING 705 727 C2H2-TYPE.
FT ZN_FING 733 755 C2H2-TYPE.
FT ZN_FING 761 783 C2H2-TYPE.
FT ZN_FING 789 811 C2H2-TYPE.
SQ SEQUENCE 818 AA; 94171 MW; 759FDB6338AD28D CRC64;

Query Match 5.5%; Score 288; DB 1; Length 818;
Best Local Similarity 20.7%; Pred. No. 4.6e-06;
Matches 131; Conservative 66; Mismatches 249; Indels 188; Gaps 18;

OY 368 GGNHKKRCAKVPESDSKLOHLBSHTGERPRKCNVCGNRFTTRGNLKVPHRRKPYH 427
DB 254 GKPYKCNCEGKAFQNSNLTSRRISHSEKPYKSCGKFTVRSMLTIHQVHTGKPY 313
OY 428 VQNNPVPVPEHLVDVITSSGLPYGMSVPEKAEENATPGGVERKPLVASTALATES 487
DB 314 KCEGCKGVFRRHNSLTLAHRRIHTGK--PYKCN----- 345
OY 488 LTLTTSAGTATAPGLPAFNKVFVLKAVEPKKAKADENTPGSEGSATSGVAESSTATLM- 546
DB 346 -----CGKAFRRGNSNLTHQI 363
OY 547 -----QLSKMTSLSPSALLTNMK-STGSPFLPL--CARALGASPSKLOOLV 594
DB 364 TGEKPFKCNCGKLFTON---SHLISHRRIHTEKPYKCNCEGKAF--SVRSSLAHQTI 418
OY 595 EKIDROGAVAVTASASGAPTTASAPASSASSGPNOCVICTRLVSCRALRLHVGOGGE 654
DB 419 HTGEEK-----PYKCNCEGKAFRRNSYIGRRHRRVHTGE 450
OY 655 RPKCKVCVGAFTSGNLRAHFVGHKASPARAONSCPIQCKKETTNAVTLQOHVHMLG 714
DB 451 KPYKCNCGKAFSMHSNLTATHQVHTGKPFK---CNCKSVKFTQSGQLANHRRIHTG- 505
OY 715 QIINGGTALEBGGAAOENSGOSTVSGAGSPPOOQSOPSEEEISEEEDEEED 774
DB 506 -----EKPYKCNCGKAFSV-----RSSLITTHQAIHSEKPYKCIEGGKS 545
```

```
OY 775 VTDEDSLGRSGESGGEKALIVRGDSEASGAEEVGTVAATAAT----- 818
DB 546 FTQKSHLRSHHGHIHSGEPRYK-----NECGKVFAGTQSOLARHMRVHTGKPYK 594
OY 819 ---AGKEMDSNEKTTQGSLLPPEPPDSLDQPPMEQSGSVLGK--EEGKPERSSSP 873
DB 595 CNDCGRAFSRDSRLTFHQAIHTGKPYKCEGKVFRRHNSYLTATHRIHTGKPYKCNCG 654
OY 874 ASALTPEGEATSVTLVEELSLDEAKRKFESESSRKACEVCGCAFPSQALFEHQTHPK 933
DB 655 GKAFSMHSNLTTHKVI-----HTGKPYK--CNQCGKVFQNSHLANHQRTHTG 701
OY 934 EGPLFTCVFCROGFLEATLKKMLLAHHVOPE 967
DB 702 EKP-YRCNCGKAFSVKSSLTTHQAI-HTGKPY 733

RESULT 21
ZFHL_DROME STANDARD; PRT: 1060 AA.
AC P2816;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Zinc-finger protein 1 (zinc-finger homeodomain protein 1).
GN ZFH-1.
OS Drosophila melanogaster (fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92001539; PubMed=1680376;
RA Fortini M.E., Lai Z., Rubin G.M.;
RT "The Drosophila zfh-1 and zfh-2 genes encode novel proteins
containing both zinc-finger and homeodomain motifs.";
RL Mech. Dev. 34:113-122(1991).
CC -1- FUNCTION: INVOLVED IN THE DEVELOPMENT OF THE EMBRYONIC CENTRAL
NERVOUS SYSTEM, EMBRYONIC MESODERM AND ADULT MUSCULATURE.
CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -1- TISSUE SPECIFICITY: MESODERM AND MESODERMALLY-DERIVED STRUCTURES
IN THE EMBRYO INCLUDING THE DORSAL VESSEL, SUPPORT CELLS OF THE
GONADS, AND SEGMENT-SPECIFIC ARRAYS OF ADULT MUSCLE PRECURSOR.
CC ALSO IDENTIFIED IN MOTOR NEURONS OF DEVELOPING CNS.
CC -1- SIMILARITY: BELONGS TO DELTA-EF1/ZFH-1 FAMILY OF TWO-HANDED ZINC
FINGER/HOMEODOMAIN PROTEINS.
CC -1- SIMILARITY: CONTAINS 1 HOMEODOMAIN.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@sib-sib.ch).
CC
DB EMBL: M63449; AAA29050.1; -.
DB PIR: S27816; S27816.
DB PIR: S33641; S33641.
DB HSSP: P08153; 12FD.
DB TRANSFAC: T00919; -.
DB FlyBase: FBgn0004606; zfh1.
DB InterPro: IPR001356; Homeobox.
DB InterPro: IPR000822; znf_C2H2.
DB Pfam: PF00046; homeobox; 1.
DB Pfam: PF00096; zf-C2H2; 9.
DB PRINTS: PR00048; ZINC_FINGER.
DB PRODOM: PD000003; znf_C2H2; 2.
DB PRODOM: PD000010; Homeobox; 1.
DB SMART: SM00389; HOX; 1.
DB SMART: SM00355; znf_C2H2; 9.
DB PROSITE: PS00027; HOMEODOMAIN_1; 1.
```


DR Pfam: PF01352; KRAB; 1.
 DR PRINTS: PR00048; ZINC_FINGER.
 DR ProDom: PD000003; ZnF_C2H2; 17.
 DR SMART: SM00349; KRAB; 1.
 DR SMART: SM00355; ZnF_C2H2; 33.
 DR PROSITE: PS00805; KRAB; 1.
 DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; 33.
 DR PROSITE: PS00157; ZINC_FINGER_C2H2_2; 34.
 KW Transcription regulation; DNA-binding; Zinc-finger; Metal-binding;
 KW Nuclear protein; Repeat.
 FT DOMAIN 7 75
 FT ZINC_FINGERS.
 FT ZINC_FINGER (DEGENERATE).
 FT ZN_FING 201 1147
 FT ZN_FING 201 223
 FT ZN_FING 229 251
 FT ZN_FING 257 279
 FT ZN_FING 285 307
 FT ZN_FING 313 335
 FT ZN_FING 341 363
 FT ZN_FING 369 391
 FT ZN_FING 397 419
 FT ZN_FING 425 447
 FT ZN_FING 453 475
 FT ZN_FING 481 503
 FT ZN_FING 509 531
 FT ZN_FING 537 559
 FT ZN_FING 565 587
 FT ZN_FING 593 615
 FT ZN_FING 621 643
 FT ZN_FING 649 671
 FT ZN_FING 677 699
 FT ZN_FING 705 727
 FT ZN_FING 733 755
 FT ZN_FING 761 783
 FT ZN_FING 789 811
 FT ZN_FING 817 839
 FT ZN_FING 845 867
 FT ZN_FING 873 895
 FT ZN_FING 901 923
 FT ZN_FING 929 951
 FT ZN_FING 957 979
 FT ZN_FING 985 1007
 FT ZN_FING 1013 1035
 FT ZN_FING 1041 1063
 FT ZN_FING 1069 1091
 FT ZN_FING 1097 1119
 FT ZN_FING 1125 1147
 FT ZN_FING 1167 AA; 134352 MW; E2184DF23BD035E9 CRC64;
 SQ SEQUENCE
 Query Match 5.3%; Score 278.5; DB 1; Length 1167;
 Best Local Similarity 21.9%; Pred. NO. 1.7e-05;
 Matches 143; Conservative 76; Mismatches 266; Indels 167; Gaps 24;

Db 703 KPYKCECGKTFESKVSSTLTTHKAIHAGEKPYKCECGKAFNSKFSILTKHKYIHTEKPYK 762
 QY 687 AONSCLPCKKFTNAVTLQOYHVMILGGIPIKNGTALPREGGAAGENGSDSTVSGACSF 746
 Db 763 ---CEECGKAYKWPSTLSYHKHIHTG----- 785
 QY 747 PQQSOQPSPEELSEEEDEEEDVTDDESLAGSGSGGEKA---ISVGDSEEA 803
 Db 786 ----EKPYKCECGKSGMSILTKHEYIHGEKPYKCECG--KAFSLSVSKHKKT 838
 QY 804 SGAEVEGYVAAATAGKEMDSNEKTTQSSLP PPPPSLDQPQMEGSSGVLGKE- 862
 Db 839 HAGEK----FYKCEACGKAYNTFSILTKHYIHTGEKPYKCECGKAFNMSSNLMHKKI 894
 QY 863 -EGKRP-----ERSSSPASALTPGCAATSVYLVELSLQEAHKKEGESSRRACEVCG 915
 Db 895 HTGETPYKCECGDKAFSPSSLT-EHKAT-----HAGEKPYK--CECG 935
 QY 916 QAFPSOALAEHOKTHPKEGPLFTGCVGROGFLERATLKKHMLLAHQVOFP 967
 Db 936 KAFSPSLRTEHKATHAGEEP-YKCECGKAFNMSSNLMHKKI-HTEKPY 985
 RESULT 23
 Y222.HUMAN STANDARD: PRT: 1163 AA.
 AC Q92618;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Hypothetical zinc finger protein KIAA0222.
 GN KIAA0222
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Bone marrow;
 RX MEDLINE=97191544; PubMed=9039502;
 RA Nagase T., Seki N., Ishikawa K.-I., Ohira M., Kawarabayasi Y.,
 Ohara O., Tanaka A., Kotani H., Miyajima N., Nomura N.;
 RT "Prediction of the coding sequences of unidentified human genes. VI.
 The coding sequences of 80 new genes (KIAA0201-KIAA0280) deduced by
 RT analysis of cDNA clones from cell line KG-1 and brain.";
 RL DNA Res. 3:321-329(1996).
 CC -1- FUNCTION: MAY FUNCTION AS A TRANSCRIPTION FACTOR.
 CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).
 CC -1- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-
 FINGER PROTEINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb.ch/announce/>
 CC or send an email to license@isb.slb.ch).
 CC -----
 DR EMBL: D86975; BAAL3211.1; -;
 DR InterPro: IPR000822; ZnF_C2H2.
 DR Pfam: PF000096; Zf-C2H2; 10.
 DR PRINTS: PR00048; ZINC_FINGER.
 DR ProDom: PD000003; ZnF_C2H2; 1.
 DR SMART: SM00355; ZnF_C2H2; 10.
 DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; 7.
 DR PROSITE: PS00157; ZINC_FINGER_C2H2_2; 7.
 KW Hypothetical protein; Transcription regulation; DNA-binding;
 KW Zinc-finger; Metal-binding; Nuclear protein; Repeat.
 FT DOMAIN 34 56
 FT ZN_FING 34 1120
 FT ZN_FING 62 84
 FT ZN_FING C2H2-TYPE (ATYPICAL).

FT	2N_FING	174	197	C2H2-TYPE.	
FT	2N_FING	200	223	C2H2-TYPE (ATYPICAL).	
FT	2N_FING	248	270	C2H2-TYPE.	
FT	2N_FING	276	298	C2H2-TYPE.	
FT	2N_FING	335	357	C2H2-TYPE.	
FT	2N_FING	515	537	C2H2-TYPE.	
FT	2N_FING	760	783	C2H2-TYPE (ATYPICAL).	
FT	2N_FING	1098	1120	C2H2-TYPE.	
SO	SEQUENCE	1163 AA; 124288 MW; 9D6DDDED2D7023644 CRC64;			
Query Match					
Best Local Similarity 5.3%; Score 278; DB 1; Length 1163;					
Matches 161; Conservative 75; Mismatches 266; Indels 264; Gaps 28;					
OY	362	GPLEKPGGR-----	-HKRCFAKVFSGDSALQIHLRSHTEGRRPYKCNOCNPFET	410	
DB	14	GSPTTAGRHEVDGDKATCTCCICGSPFOSSLSQHMRTKTEKPYCPYCDHRASQ		73	
OY	411	RCNLKVFHRRREKYPVHOMNHPVPEHLIDYVITSSGLPYGNSVPPEKAEAEATPGGCV		470	
DB	74	KGNLKHIRSHRGITLLQGHPEPAGEARPLGEMRASBGL-----	-DACASP----	117	
OY	471	EKKPLVASTALSTESILTLSTAGTAPGLPAPNKFVLKMAVEPKKAD-----	EN 524		
DB	118	-----FKSASACNRL-----	NGASQADGARYLN-----	GASQADSGRVLRS 155	
OY	525	TPPGSGSATSIVAVESSTATLMOLSKMTSLPBMALTNH-FKSTGSEPLPLCARALGAS		583	
DB	156	SKKGAESA---CAPGEAKAVNOCFSCKSOFRKKDLHIVHQAHPKCRCLSYA----		208	
OY	584	PSETSKLOOLVEKIDROGAVAVTSAASGAPPTSADAPSSSASGPNQCVYCLRYLSCPRA		643	
DB	209	---TLREESLSHIERD---HITNAGPSGGEACVENGKPELSPGEPCVCGGAFSQTWF		262	
OY	644	LRHYNQHGEBRPPCKVCYCGAFTSGNIRAFVGH-----	-KASPARAON--	689	
DB	263	LKAHMKHRSFPHGHCICGRFRPEWFLKNMKAHGPTGSKNRPKSELDPATINNIV		322	
OY	690	-----SCPIQOKKFTNAVTLQOHVHMLGGQIPNGCTALPEGGGAQENGSE		736	
DB	323	QEEVYVAGSLVEYCAACGNIPLTINDLSLNAHNAIH--RVEASRTAPAEAGAGPSDTK		380	
OY	737	QSTV-----	SGAG-SFP-----	QOQSQPSPEELSEEE-----	765
DB	381	QEPLOCLNLRPSAAGDSCTGQAGRVAELDPVNSYQAWOLATRGVAEPAYELIKYGAMD		440	
OY	766	-----	BEDEEE-----	EEDVTDSDSLAGRG-----	SESGGEKATSVRGDSEASG 805
DB	441	EALAGVAFPKDRREVLYVSOEKRRREDAPARAQGPRRRASGPDPAAPAGHLDPRSAAR		500	
OY	806	AEEEVGTVAATAAG-----		KEMDSNEKTTQOSS 834	
DB	501	PNNR-----	AAATTGQCKSSSECFECGKIFRTYHOMVLSRVRARERSDGDRAARAR		555
OY	835	LPPPPPSLDOPPMBOGS-----		SGVLGGR-EGRKPER----	869
DB	556	CSLSGDSASQPS--SPGSACAADSPGSLADEAEDSGEGGADPEAPAGGPRRCFS		613	
OY	870	-----		SSPASALTPBEATSVTLVEELSDAMKE-----	901
DB	614	EEVTSLELSSGDOSHKMKDNASEBDTGESKAGIAASVLSILENSSRSTSRQEOHFRFSMDL		673	
OY	902	-----		PGESSSRKACEVCGOAFPSOALAEHOHTHPKE 934	
DB	674	KMPAFHPKQEVVPFGD-----		GVEFPESSTG-ABGQTGHPAE 708	
RESULT 24					
ID	2133_HUMAN	STANDARD;	PRT;	654 AA.	
AC	P52736; Q9H443; Q9BUV2;				
DT	01-OCT-1996 (rel. 34, Last sequence update)				

DT	15-JUN-2002 (Rel. 41, Last annotation update)
DE	Zinc finger protein 133.
GN	ZNF133.
OS	Homo sapiens (Human).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX	NCBI_TaxId=9606;
RN	[1]
RC	SEQUENCE FROM N.A.
RP	TISSUE=Insulinoma;
RX	MEDLINE=95377390; PubMed=7649249;
RA	Vissing H., Meyer W.-K., Aagaard L., Tommerup N., Thiesen H.-J.;
RT	"Repression of transcriptional activity by heterologous KRAB domains
RT	present in zinc finger proteins.";
RL	FEBS Lett. 369:153-157(1995).
RN	[2]
RP	SEQUENCE FROM N.A.
RX	MEDLINE=21638749; PubMed=11780052;
RA	Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R.,
RA	Jones M., Stavrides G., Almeida J.P., Babbage A.K., Bagguley C.L.,
RA	Bailey J., Barlow K.F., Bates K.N., Bead L.M., Beare D.M.,
RA	Beasley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J.,
RA	Buck D., Burrill W.D., Butler A.P., Carder C., Carter N.P.,
RA	Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M.,
RA	Clegg S., Cobley V.E., Collier R.E., Connor R.E., Corby N.R.,
RA	Coulson A., Coville G.J., Deadman R., Dhani P.D., Dunn M.,
RA	Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,
RA	Graham D.V., Griffiths C., Griffiths M.N.D., Gilliam R., Hall R.E.,
RA	Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J.,
RA	Huckle E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D.,
RA	Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,
RA	Lhevassialho M.H., Levesha M.A., Lloyd C., Lloyd D.M., Lovell J.D.,
RA	Marsh V.L., Martin S.L., McConachie L.J., McIay K., McMurray A.A.,
RA	Milne S.A., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T.,
RA	Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.T.,
RA	Phillimore B.J.C.T., Prathalingam S.R., Plumb R.W., Ramsay H.,
RA	Rice C.M., Ross M.T., Scott C.E., Sehra H.K., Showkhen R., Sims S.,
RA	Skue C.D., Smith M.L., Soderlund C., Steward C.A., Sulston J.E.,
RA	Swann R.M., Sycamore N., Taylor R., Tee L., Thomas D.W., Thorpe A.,
RA	Tracey A., Tromans A.C., Vaudin M., Wall M., Wallis J.M.,
RA	Whitehead S.L., Whitaker P., Willey D.L., Williams L., Williams S.A.,
RA	Whitling L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,
RA	Rogers J.;
RL	"The DNA sequence and comparative analysis of human chromosome 20.";
RN	Nature 414:865-871(2001).
RN	[3]
RP	SEQUENCE FROM N.A.
RC	TISSUE=Lung;
RA	Strausberg R.;
RL	Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
CC	- FUNCTION: MAY BE INVOLVED IN TRANSCRIPTIONAL REGULATION AS A
CC	REPRESSOR.
CC	- SUBCELLULAR LOCATION: Nuclear (Potential).
CC	- TISSUE SPECIFICITY: SEEMS UBICUITOUS. SEEN IN THE HEART, BRAIN,
CC	PLACENTA, LUNG, LIVER, SKELETAL MUSCLE, KIDNEY, AND PANCREAS.
CC	- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-
CC	FINGER PROTEINS.
CC	- SIMILARITY: CONTAINS 1 KRAB DOMAIN.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC	the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC	or send an email to license@isb-sib.ch).
CC	-----
DR	EMBL; U09366; AAC50260.1; -
DR	EMBL; AL049646; CAC15508.1; -
DR	EMBL; BC001887; AA010887.1; -
DR	HSSP; P08045; 1ZNF.
DR	TRANSFAC; T04992; -
DR	GeneW; HGNC:12917; ZNF133.


```
DR MIM: 604075: -
DR Interpro: IPR001909: KRAB.
DR Interpro: IPR000822: ZnF_C2H2.
DR Pfam: PF00096: ZF-C2H2; 30.
DR Pfam: PF01352: KRAB; 1.
DR PRINTS: PR00048: ZINCINGER.
DR Prodom: PD000003: ZnF_C2H2; 5.
DR SMART: SM00349: KRAB; 1.
DR SMART: SM00355: ZnF_C2H2; 15.
DR PROSITE: PS00805: KRAB; 1.
DR PROSITE: PS00028: ZINC_FINGER_C2H2_1; 14.
DR PROSITE: PS00157: ZINC_FINGER_C2H2_2; 15.
DR Transcription regulation: Zinc-finger; DNA-binding; Metal-binding;
KW Nuclear protein; Repeat; Repressor.
FT DOMAIN 1 772 KRAB.
FT ZINC_FINGERS.
FT ZN_FING 214 631
FT ZN_FING 214 236 C2H2-TYPE.
FT ZN_FING 242 264 C2H2-TYPE.
FT ZN_FING 270 292 C2H2-TYPE.
FT ZN_FING 298 320 C2H2-TYPE.
FT ZN_FING 326 348 C2H2-TYPE.
FT ZN_FING 354 376 C2H2-TYPE.
FT ZN_FING 382 404 C2H2-TYPE.
FT ZN_FING 410 432 C2H2-TYPE.
FT ZN_FING 438 460 C2H2-TYPE.
FT ZN_FING 466 488 C2H2-TYPE.
FT ZN_FING 494 516 C2H2-TYPE.
FT ZN_FING 522 544 C2H2-TYPE.
FT ZN_FING 550 572 C2H2-TYPE.
FT ZN_FING 578 600 C2H2-TYPE.
FT ZN_FING 606 631 C2H2-TYPE (ATYPICAL).
FT CONFLICT 73 73 MISSING (IN REF. 3).
FT CONFLICT 193 193 T->S (IN REF. 2).
SQ SEQUENCE 654 AA: 73402 MW: 2042674532024228 CRC64;

Query Match
Best local similarity 5.2%; Score 277; DB 1; Length 654;
Matches 177; Conservative 67; Mismatches 273; Indels 352; Gaps 34;

152 PESTPAPPAPP---PPPPPGVSGHLNIPILILELRLVLOQRQIHQMOTETQICRQ----- 204
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db PATCPAPPEPEPELYLDPECPPEPSS-----QKPEPMQHVLCNHPWI 107

205 VLLIGSGQVYGAPASPELDEGTGASTKP-----LLPEPPIKPVQTSKTL 252
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db FTCLCAGNIOGPBGPEGOEKQOASERPSWSDQABEGPEGAMPLFGRTK---KITL 163

253 ASSSSSSSSSGAETPKQAFPHLYHPLGSOHPFSAGVGSRSHKPTTPAPSPALPGSTDOLI 312
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db GARS-----RPPQR-----QPVSSRN-----GLRGVELEASPAQT-GNPETDKLL 203

313 -----ASPHLAPSTGTLLAOCAGANGLEATASPGLLKPKNGSGELSYGE 359
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db KRLEVLFGTVNCGECGGSFSKMTNLSHQ-----RINSGEKPY-- 242

204 VMG-----PLEKPGGRHK-----CRCAKYFGSDSALQIHLRSHNGERYKCNV 403
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db VCGVCEGFGFSLKSLAKQKHAHSGEKPIYCREGGRNKRSTLLIHEHTHSGEKPYMCSE 302

243 CGNFTTGRNLKVHHRHREKYPHYQMNPHVPEHLDYVITSSGLPYGMSVPEKAEDEEA 463
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db CGRGFSQKSNLIHQRTHSGEKPY-----YVCREGCKGFSQKSAVYVNRHQT- 347

464 ATPGGYERKLVASTYALSTATESLTLSTASGTATAFGLPAPKVEYLMKAVEPKNKADE 523
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db -----HLEEKTYVCDGCGFSDRSNLIS-----HQ 373

524 NTPGREGSATS--GVASSSTATLMQSLKMTSLPSSWALLTNHFKTSGSPFLPLCARALG 581
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db KTHSGEKPYAKCEGRCRQRQRTLV-----NH-QRTHSKERPIYCGVCG 416

582 ASPSETSKLOOLVEKIDRGAVAVTSAASGAPPTTAPAPSSSASGPNQCVICLRVLSCP 641
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
```

```
Db 417 HSFQSNSTL-----ISHR-----RTHNGEKPIYGVCGRGFSLK 450
Oy 642 RALRLHYGQHGGERPFKKVCGRATSTRCNLRAHFQHKASPARAONSCICQKKFTNA 701
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 451 SHLNRRHONHSGEKPIYCKDKGRGFSQOQSNL-----TRHQRTHSGEKPMVCGEGRGFSQK 506
Oy 702 VTLQAHVMHIGGQIPNCGTALPEGGAQOENGSQSTVSGAGSFPOQSOQSPPEELSL 761
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 507 SNLYAHQRTHSG-----ERYVCREG----- 528
Oy 762 EEEDEDEDEEDVDDEDLAGRGSESGEKAISVGRDSEASGAEVEGTVAATAAGK 821
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 529 -----RGFSHQ-----AGLRHK 541
Oy 822 EMDSNEKTTQSSLPPEPPPSLDQPPMEGSSGVLCGEDEGKPERSSSPASALPREG 881
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 542 RKHSREK-----PYMCRQCGLG-FGNK-----SALITRK 569
Oy 882 EATSVTLVEELSLQAMRKPEGESSRKACEVCGQAFPSQAALBEHOKTHPEKGPFTCV 941
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 570 RAHS-----EKKP-----CYCREGGGFLQKSHLTHQMTHTGKRP-YVCK 609
Oy 942 FCROGFLEATLKKH--MLAHQ--VQP 966
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 610 TCGRGFSLSKSHLSRHKRTTSVHHRLPVOP 638

RESULT 25
2268_HUMAN STANDARD: PRT; 947 AA.
AC 014587; 09BZJ9; 096RH4.
DT 16-OCT-2001 (Rel. 40, last sequence update)
DT 15-JUN-2002 (Rel. 41, last annotation update)
DE Zinc finger protein 268 (Zinc finger protein HZF3).
GN ZNF268.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
RC TISSUE=Embryo;
RX MEDLINE=21210976; PubMed=11311945;
RA Gou D.M., Sun Y., Gao L., Chow L.M.C., Huang J., Feng Y.D.,
   Jiang D.H., Li W.X.;
RT "Cloning and characterization of a novel Kruppel-like zinc finger
   gene, ZNF268, expressed in early human embryo.";
RL Biochim. Biophys. Acta 1518:306-310(2001).
RN [2]
RP SEQUENCE OF 549-947 FROM N.A.
RX MEDLINE=95169271; PubMed=7865130;
RA Airlink M., Aveskogh M., Hellman L.;
RT "Isolation of cDNA clones for 42 different Kruppel-related zinc finger
   RT proteins expressed in the human monoblast cell line U-937.";
RL DNA Cell Biol. 14:125-136(1995).
CC -|- FUNCTION: MAY FUNCTION AS A TRANSCRIPTION FACTOR.
CC -|- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -|- ALTERNATIVE PRODUCTS: 2 isoforms; A/ZNF268A (shown here) and
   B/ZNF268B; are produced by alternative splicing.
CC -|- DEVELOPMENTAL STAGE: Highly expressed in 3- to 5-week-old embryos.
CC -|- SIMILARITY: BELONGS TO THE KRUPPEL FAMILY OF C2H2-TYPE ZINC-
   FINGER PROTEINS.
CC -|- SIMILARITY: CONTAINS 1 KRAB DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb.ch).
```



```
FT DOMAIN 21 239 ZINC-FINGERS 1.
FT DOMAIN 421 434 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT DOMAIN 724 803 ZINC-FINGERS 2.
FT DOMAIN 877 928 ASP/GLU-RICH (ACIDIC).
FT ZN_FING 21 44 C2H2-TYPE.
FT ZN_FING 75 97 C2H2-TYPE.
FT ZN_FING 103 125 C2H2-TYPE.
FT ZN_FING 131 154 C2H2-TYPE.
FT ZN_FING 160 182 C2H2-TYPE.
FT ZN_FING 188 210 C2H2-TYPE.
FT ZN_FING 217 239 C2H2-TYPE.
FT ZN_FING 724 746 C2H2-TYPE.
FT ZN_FING 752 775 C2H2-TYPE.
FT ZN_FING 781 803 C2H2-TYPE.
FT SITE 553 557 CTBP-BINDING MOTIF 2 (BY SIMILARITY).
FT SITE 584 588 CTBP-BINDING MOTIF 1 (BY SIMILARITY).
SQ SEQUENCE 1042 AA: 116847 MW: 8DEDF164F536D2PE CRC64;

Query Match 5.2%; Score 276; DB 1; Length 1042;
Best Local Similarity 18.7%; Pred. No. 2e-05;
Matches 176; Conservative 87; Mismatches 245; Indels 432; Gaps 34;

OY 368 GGRHAKRFCAKVGSDSALQIHLDS-HTGERPKVCNVCGRNRTTGNLKVHFRHREKYP 426
D 128 GKHYECENCAKVFDPNLRHRSQHVGAHACPECKTEFTSSGLKQHKHHSVKP 187
OY 427 -----HYQMN----- 431
D 188 FICEVCHKSTQFSNLCRHKRMHADCRFOIKCKDCGQMFSTTSLKNHRRFCEGKNHPAA 247
OY 432 -----PHRPEHLDYVITSSGLPYGMSVPEPKA---EEEA 464
D 248 GGFEGGICISLGPTRAMKTSNVNKNHNPGLADYGTGNH-PAGLTFPTAPGSPFSPGL 306
OY 465 TRGGVGRKPLVAST---TALSAE-----SLTLSTAGATADGCPAENKFLVM--- 512
D 307 FPGGLYHRPLIPASPPVKGLSTFQSNKCGSPRLTHPOLPILPATQDILKALSKHPYGDN 366
OY 513 KAVE--PKNADE-----NTPGSE-----GSAISGAESSTALIMQ 547
D 367 KPELLPERSESEERLEKISQSSSSLDVSTPGSDLETTSGSDLESDEKKECKE 426
OY 548 LSKL-----MTSLPMTLTN-----HFK 566
D 427 NGMFMKDKVPLQGLASTNNKEHNHNSVFSASVEBQSAVSGAVNDSIKALASIAETFG 486
OY 567 STG-----SFPLP-----LCARALGASPSSTKIQ 591
D 487 STGLVGLQDKKVGALPYRSMPLPFPPAFSGSMYPPRDRLRSLPLKMEPOSPEVKLIQ 546
OY 592 Q-----LVERIDQGAVAVTSASGAPTTSAAP-----SSASG----- 627
D 547 KGSSESPFDLTTRKDKERPLTSGPSPGTSPATSDQPLDLSMGSRCRASGTLKLEPRKN 606
OY 628 ----- 627
D 607 HVEGEKKGSNMDTSPSSGSLQHARPTPEFMDPIYVEKRLTLDPLALEKYLRLPSRGF 666
OY 628 -----PN-----QC 631
D 667 LHPHOMSAIENMAEKLESFALKPEASELLQVPSMFSFAPRNTPLPENLIRKGERYTC 726
OY 727 RYCGKLT--PRSANLTRLRLTHHTGEOPYRCKYCDRFSISSNQRHNRIHNKEKP--- 781
OY 688 QNSCPLQCKFTNAVTLQOAHVMMHGGQIPNGGTALP-----ECGGAA----- 730
D 782 --KCHLDCRCGQGTNIDRHLKKHNGMNSGATSSPHSELSEAGALLDKEDAVFTPEIR 839
OY 731 -----QENGSE-----OSTVSGAGSPPOQSOQPPSEBELSEEEDE---BEEDVT 776
D 840 NFIGNSHNGSQSPRNMERMGNS-HFKDKKALATSONSDLLDDEEVEDEVLLDEDEDD-- 896
```

```
OY 777 DEDSLAGRSGESGGEKAISVRGDSEASGAEVEVGTVAATAAG----- 820
D 897 --NDIPKPKRELG-----VTRUDEEIPEDDYREAGLEMSCKSPVKRYKREDYKSGLSAL 950
OY 821 -----KEMDSNEKT-TQSSSLPPPPPSLDQPPQMEQSSGVLAGKEEGKP 867
D 951 DHIRHFTDSLKRMEENQYTDALSLSSISSHVPELKTQLHKK-----SKQAVAM 1002
OY 868 ERSSSPASALTPEGEATSVTLVEELSLQELAMRKREPESSS 907
D 1003 MLSLSDKDSLHPTSHSSS-----NWMHSMARMAAESSA 1035

RESULT 27
ID ZN43_HUMAN STANDARD; PRT; 803 AA.
AC P17038; P28160; Q96DG1;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Zinc finger protein 43 (Zinc protein HTF6) (Zinc finger protein KOX27).
GN ZNF43 OR ZNF39 OR KOX27.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-T-cell;
RX MEDLINE=91279444; PubMed=1711675;
RA Lovering R., Trowsdale J.;
RT "A gene encoding 22 highly related zinc fingers is expressed in
RL lymphoid cell lines.";
RN Nucleic Acids Res. 19:2921-2927 (1991).
[2]
RP SEQUENCE FROM N.A.
RC TISSUE-Lymph;
RA Strausberg R.;
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 38-190 FROM N.A.
RX MEDLINE=91219421; PubMed=2023909;
RA Bellefroid E.J., Poncelet D.A., Lecocq P.J., Revelant O.,
RA Martial J.A.;
RT "The evolutionarily conserved Kruppel-associated box domain defines a
RL subfamily of eukaryotic multifingered proteins.";
RN Proc. Natl. Acad. Sci. U.S.A. 88:3608-3612 (1991).
[4]
RP SEQUENCE OF 476-531 FROM N.A.
RC TISSUE-Lymphoid;
RX MEDLINE=91145339; PubMed=2288909;
RA Thiesen H.-J.;
RT "Multiple genes encoding zinc finger domains are expressed in human T
RL cells.";
RN New Biol. 2:363-374 (1990).
CC - FUNCTION: MAY BE INVOLVED IN TRANSCRIPTIONAL REGULATION.
CC - SUBCELLULAR LOCATION: Nuclear (Probable).
CC - TISSUE SPECIFICITY: T AND B CELL LINES.
CC - SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-
CC FINGER PROTEINS.
CC - SIMILARITY: CONTAINS 1 KRAB DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X59244; CAA41932.1; -.
```


CR EMBL: BC006528; AAH06528.1; -.
 CR EMBL: M61869; AAA58674.1; -.
 CR EMBL: X52358; CAA36584.1; -.
 CR PIR: S26823; S26823.
 CR PIR: D39384; D39384.
 CR PIR: S10416; S10416.
 CR HSSP: P08048; 7ZNF.
 CR TRANSFAC: T04986; -.
 CR Genew: HGNC:13109; ZNF43.
 CR MIM: 603972; -.
 DR InterPro: IPR001909; KRA-B.
 DR InterPro: IPR000822; Znf_C2H2.
 DR Pfam: PF00096; zf-C2H2; 21.
 DR Pfam: PF01352; KRA-B; 1.
 DR PRINTS: PRO00048; ZINCFINGER.
 DR PRODOM: P000003; Znf_C2H2; 16.
 DR SMART: SM00349; KRA-B; 1.
 DR SMART: SM00355; znf_C2H2; 21.
 DR PROSITE: PS50805; KRA-B; 1.
 DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; 19.
 DR PROSITE: PS50157; ZINC_FINGER_C2H2_2; 22.
 KW Transcription regulation; DNA-binding; Zinc-finger; Metal-binding;
 FT Nuclear protein; Repeat.
 FT DOMAIN 1 69 KRA-B, ZINC_FINGERS.
 FT ZN_FING 168 778 ZINC_FINGERS.
 FT ZN_FING 168 190 C2H2-TYPE.
 FT ZN_FING 196 218 C2H2-TYPE (DEGENERATE).
 FT ZN_FING 224 246 C2H2-TYPE (DEGENERATE).
 FT ZN_FING 252 274 C2H2-TYPE (DEGENERATE).
 FT ZN_FING 280 302 C2H2-TYPE.
 FT ZN_FING 308 330 C2H2-TYPE.
 FT ZN_FING 336 358 C2H2-TYPE.
 FT ZN_FING 364 386 C2H2-TYPE.
 FT ZN_FING 392 414 C2H2-TYPE.
 FT ZN_FING 420 442 C2H2-TYPE.
 FT ZN_FING 448 470 C2H2-TYPE.
 FT ZN_FING 476 498 C2H2-TYPE.
 FT ZN_FING 504 526 C2H2-TYPE.
 FT ZN_FING 532 554 C2H2-TYPE.
 FT ZN_FING 560 582 C2H2-TYPE.
 FT ZN_FING 588 610 C2H2-TYPE.
 FT ZN_FING 616 638 C2H2-TYPE.
 FT ZN_FING 644 666 C2H2-TYPE.
 FT ZN_FING 672 694 C2H2-TYPE.
 FT ZN_FING 700 722 C2H2-TYPE.
 FT ZN_FING 728 750 C2H2-TYPE.
 FT ZN_FING 756 778 C2H2-TYPE.
 FT CONFLICT 181 181 P -> S (IN REF. 3).
 FT CONFLICT 712 712 P -> S (IN REF. 2).
 FT SEQUENCE 803 AA; 93487 MW; 88C41B44DFDE2533 CRC64;
 SO SEQUENCE
 Query Match 5.2%; Score 275; DB 1; Length 803;
 Best Local Similarity 20.1%; Pred. No. 1.7e-05;
 Matches 135; Conservative 59; Mismatches 194; Indels 282; Gaps 22;
 CR EMBL: BC006528; AAH06528.1; -.
 CR EMBL: M61869; AAA58674.1; -.
 CR EMBL: X52358; CAA36584.1; -.
 CR PIR: S26823; S26823.
 CR PIR: D39384; D39384.
 CR PIR: S10416; S10416.
 CR HSSP: P08048; 7ZNF.
 CR TRANSFAC: T04986; -.
 CR Genew: HGNC:13109; ZNF43.
 CR MIM: 603972; -.
 DR InterPro: IPR001909; KRA-B.
 DR InterPro: IPR000822; Znf_C2H2.
 DR Pfam: PF00096; zf-C2H2; 21.
 DR Pfam: PF01352; KRA-B; 1.
 DR PRINTS: PRO00048; ZINCFINGER.
 DR PRODOM: P000003; Znf_C2H2; 16.
 DR SMART: SM00349; KRA-B; 1.
 DR SMART: SM00355; znf_C2H2; 21.
 DR PROSITE: PS50805; KRA-B; 1.
 DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; 19.
 DR PROSITE: PS50157; ZINC_FINGER_C2H2_2; 22.
 KW Transcription regulation; DNA-binding; Zinc-finger; Metal-binding;
 FT Nuclear protein; Repeat.
 FT DOMAIN 1 69 KRA-B, ZINC_FINGERS.
 FT ZN_FING 168 778 ZINC_FINGERS.
 FT ZN_FING 168 190 C2H2-TYPE.
 FT ZN_FING 196 218 C2H2-TYPE (DEGENERATE).
 FT ZN_FING 224 246 C2H2-TYPE (DEGENERATE).
 FT ZN_FING 252 274 C2H2-TYPE (DEGENERATE).
 FT ZN_FING 280 302 C2H2-TYPE.
 FT ZN_FING 308 330 C2H2-TYPE.
 FT ZN_FING 336 358 C2H2-TYPE.
 FT ZN_FING 364 386 C2H2-TYPE.
 FT ZN_FING 392 414 C2H2-TYPE.
 FT ZN_FING 420 442 C2H2-TYPE.
 FT ZN_FING 448 470 C2H2-TYPE.
 FT ZN_FING 476 498 C2H2-TYPE.
 FT ZN_FING 504 526 C2H2-TYPE.
 FT ZN_FING 532 554 C2H2-TYPE.
 FT ZN_FING 560 582 C2H2-TYPE.
 FT ZN_FING 588 610 C2H2-TYPE.
 FT ZN_FING 616 638 C2H2-TYPE.
 FT ZN_FING 644 666 C2H2-TYPE.
 FT ZN_FING 672 694 C2H2-TYPE.
 FT ZN_FING 700 722 C2H2-TYPE.
 FT ZN_FING 728 750 C2H2-TYPE.
 FT ZN_FING 756 778 C2H2-TYPE.
 FT CONFLICT 181 181 P -> S (IN REF. 3).
 FT CONFLICT 712 712 P -> S (IN REF. 2).
 FT SEQUENCE 803 AA; 93487 MW; 88C41B44DFDE2533 CRC64;
 SO SEQUENCE
 Query Match 5.2%; Score 275; DB 1; Length 803;
 Best Local Similarity 20.1%; Pred. No. 1.7e-05;
 Matches 135; Conservative 59; Mismatches 194; Indels 282; Gaps 22;
 CR EMBL: BC006528; AAH06528.1; -.
 CR EMBL: M61869; AAA58674.1; -.
 CR EMBL: X52358; CAA36584.1; -.
 CR PIR: S26823; S26823.
 CR PIR: D39384; D39384.
 CR PIR: S10416; S10416.
 CR HSSP: P08048; 7ZNF.
 CR TRANSFAC: T04986; -.
 CR Genew: HGNC:13109; ZNF43.
 CR MIM: 603972; -.
 DR InterPro: IPR001909; KRA-B.
 DR InterPro: IPR000822; Znf_C2H2.
 DR Pfam: PF00096; zf-C2H2; 21.
 DR Pfam: PF01352; KRA-B; 1.
 DR PRINTS: PRO00048; ZINCFINGER.
 DR PRODOM: P000003; Znf_C2H2; 16.
 DR SMART: SM00349; KRA-B; 1.
 DR SMART: SM00355; znf_C2H2; 21.
 DR PROSITE: PS50805; KRA-B; 1.
 DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; 19.
 DR PROSITE: PS50157; ZINC_FINGER_C2H2_2; 22.
 KW Transcription regulation; DNA-binding; Zinc-finger; Metal-binding;
 FT Nuclear protein; Repeat.
 FT DOMAIN 1 69 KRA-B, ZINC_FINGERS.
 FT ZN_FING 168 778 ZINC_FINGERS.
 FT ZN_FING 168 190 C2H2-TYPE.
 FT ZN_FING 196 218 C2H2-TYPE (DEGENERATE).
 FT ZN_FING 224 246 C2H2-TYPE (DEGENERATE).
 FT ZN_FING 252 274 C2H2-TYPE (DEGENERATE).
 FT ZN_FING 280 302 C2H2-TYPE.
 FT ZN_FING 308 330 C2H2-TYPE.
 FT ZN_FING 336 358 C2H2-TYPE.
 FT ZN_FING 364 386 C2H2-TYPE.
 FT ZN_FING 392 414 C2H2-TYPE.
 FT ZN_FING 420 442 C2H2-TYPE.
 FT ZN_FING 448 470 C2H2-TYPE.
 FT ZN_FING 476 498 C2H2-TYPE.
 FT ZN_FING 504 526 C2H2-TYPE.
 FT ZN_FING 532 554 C2H2-TYPE.
 FT ZN_FING 560 582 C2H2-TYPE.
 FT ZN_FING 588 610 C2H2-TYPE.
 FT ZN_FING 616 638 C2H2-TYPE.
 FT ZN_FING 644 666 C2H2-TYPE.
 FT ZN_FING 672 694 C2H2-TYPE.
 FT ZN_FING 700 722 C2H2-TYPE.
 FT ZN_FING 728 750 C2H2-TYPE.
 FT ZN_FING 756 778 C2H2-TYPE.
 FT CONFLICT 181 181 P -> S (IN REF. 3).
 FT CONFLICT 712 712 P -> S (IN REF. 2).
 FT SEQUENCE 803 AA; 93487 MW; 88C41B44DFDE2533 CRC64;
 SO SEQUENCE
 Query Match 5.2%; Score 275; DB 1; Length 803;
 Best Local Similarity 20.1%; Pred. No. 1.7e-05;
 Matches 135; Conservative 59; Mismatches 194; Indels 282; Gaps 22;
 CR EMBL: BC006528; AAH06528.1; -.
 CR EMBL: M61869; AAA58674.1; -.
 CR EMBL: X52358; CAA36584.1; -.
 CR PIR: S26823; S26823.
 CR PIR: D39384; D39384.
 CR PIR: S10416; S1041

D	b		411	----	KILTHGEK-----	-PKCECGKAFFMPS	433
O	y		643	ALRLHAYGOGGERPRPKCKVCYGRAFTGRGINRAHFVGHKASPAARAONSCPLQCKKFTNAV		702	
D	b		434	TLTKNRIRHTGEEKPYKCEYCGAKFNOPSNLT---	HKRITHAETKPYKCECGKAFRS	489	
O	y		703	TLOOHVRHHLGGQINGGTALPEGGGAQENSGSSTVSAGSPPOOQSOOPSEEEELSE		762	
D	b		490	NLTKKKTIHI-----	-EKKPYICECGK	511	
O	y		763	EEEEDEEEEDVDTEDSLAGRGSESG-----	-GEAIAISVRGDSEASG	805	
D	b		512	AFKMSKLTLEHITHTGEEKPYKCEBGCKAFNHFSILTCHKRIHTGEEKPYKC-----		562	
O	y		806	AEEEVG---TVAATAATACKEMDSNEK-----	-TTQOSSLPDPPPPDSLODOPMEQ	852	
D	b		563	--EECCGKAFTQSNNLTTHKKIHTGEEKFYKCEBGCKAFTQSSNLTTHKKIHT-----		611	
O	y		853	GSSSVLGKEKGECGKERSSSPASALTPEEGATSVTLVEELSLOEMAKRKPEPSSRRACE		912	
D	b		612	-----GKKPKCECGKAF---NOFSLTTHKKIITHEKPKYK-----	-CE	647	
O	y		913	VCGGAFPQOAALEEHOQTTPKEGPTLTCVFCROGFERATLTKKHMLAHNVQPF-----		967	
D	b		648	ECCGKAFFKMSSTLTTRKKIHTGEEKP-YKCEBGCKAFKLSTLSLTKII-HNGEKRYKCEKC		705	
O	y		968	--APHGPONI	975		
D	b		706	GKAFFNRPSNL	715		
R	SULT		28				
I	D	Z226_HUMAN		STANDARD;	PRT;	803 AA.	
A	C	Q9NYT6; Q9NS44; Q96TE6;					
D	T	16-OCT-2001 (Rel. 40; Created)					
D	T	15-JUN-2002 (Rel. 41; Last sequence update)					
D	T	15-JUN-2002 (Rel. 41; Last annotation update)					
D	E	Zinc finger protein 226.					
O	S	GN ZNF226.					
O	S	Homo sapiens (Human).					
O	C	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
O	C	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.					
O	X	NCBI_TaxID=9606;					
R	N	[1]					
R	P	SEQUENCE FROM N.A.					
R	A	Shannon M., Branscomb E., Hauser L., Gordon L., Ashworth L.,					
R	A	Stubbs L.,					
R	T	"Differential expansion of homologous zinc-finger gene families in					
R	L	human chromosome 19q13.2 and mouse chromosome 7.",					
R	N	submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.					
R	P	[2]					
R	P	SEQUENCE FROM N.A.					
R	A	Kodolyan I V., Ge Y., Severin J., Krummel G.K., Grable L.,					
R	A	Kovlyadn E., Gordon L., Shannon M., Brower A., Olsen A.S., Smith L.M.,:					
R	T	"Sequence analysis of a lmb region in 19q13.2 containing a zinc finger					
R	T	gene cluster.";					
R	L	Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.					
R	N	[3]					
R	P	SEQUENCE FROM N.A.					
R	A	Isoqal T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,					
R	A	Nishikawa T., Nagai K., Sugano S., Shiraatori A., Suo H.,					
R	A	Wagatsuma M., Hosofit T., Raku Y., Kodaita H., Kondo H., Sugawara M.,					
R	A	Takahashi S., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,					
R	A	Marubashi S., Kimura K., Murakami K., Ishi S., Kawai Y., Saito K.,					
R	A	Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K., Masuno Y.,					
R	A	Ninomiya K., Iwayanagi T.,					
R	T	"NDO human cDNA sequencing project."					
R	L	Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.					
R	N	[4]					
R	P	SEQUENCE FROM N.A.					
R	P	TISSUE=Muscle;					
R	A	Straussberg R.,:					

Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
-1- FUNCTION: MAY FUNCTION AS A TRANSCRIPTION FACTOR.
-1- SUBCELLULAR LOCATION: Nuclear (Probable).
-1- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-FINGER PROTEINS.
-1- SIMILARITY: CONTAINS 1 KRAB DOMAIN.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: AF228418; AAF34786.1; -
DR EMBL: AC074331; AAF88103.1; -
DR EMBL: AK023091; BAB14398.1; -
DR EMBL: BC024197; AAH24197.1; -
DR HSSP: P08047; 1SP2.
DR Genev: HGNC:13019; ZNF226.
DR InterPro: IPR001909; KRAB.
DR InterPro: IPR008522; Znf_C2H2.
DR Pfam: PF00096; zf_C2H2; 18.
DR Pfam: PF01352; KRAB; 1.
DR PRINTS: PR000048; ZNCFINGER.
DR PRODOM: PD000003; Znf_C2H2; 15.
DR SMART: SM00349; KRAB; 1.
DR SMART: SM00355; Znf_C2H2; 18.
DR PROSITE: PS50805; KRAB; 1.
DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; 17.
DR PROSITE: PS50157; ZINC_FINGER_C2H2_2; 19.
DR Transcription regulation; DNA-binding; Zinc-finger; Metal-binding;
KW Nuclear protein.

FT DOMAIN 8 78 KRAB.
FT ZINC_FINGERS.
FT ZINC_FINGERS (DEGENERATE).
FT ZN_FING 252 274 C2H2-TYPE
FT ZN_FING 307 329 C2H2-TYPE
FT ZN_FING 335 357 C2H2-TYPE
FT ZN_FING 363 385 C2H2-TYPE
FT ZN_FING 391 413 C2H2-TYPE
FT ZN_FING 419 441 C2H2-TYPE
FT ZN_FING 447 469 C2H2-TYPE
FT ZN_FING 475 497 C2H2-TYPE
FT ZN_FING 503 525 C2H2-TYPE
FT ZN_FING 531 553 C2H2-TYPE
FT ZN_FING 559 581 C2H2-TYPE
FT ZN_FING 587 609 C2H2-TYPE
FT ZN_FING 615 637 C2H2-TYPE
FT ZN_FING 643 665 C2H2-TYPE
FT ZN_FING 671 693 C2H2-TYPE
FT ZN_FING 699 721 C2H2-TYPE
FT ZN_FING 727 749 C2H2-TYPE
FT ZN_FING 755 777 C2H2-TYPE
FT ZN_FING 777 799 C2H2-TYPE
FT CONFLICT 195 195 K -> E (IN REF. 1).
SQ SEQUENCE 803 AA; 91920 MW; 256DE75551A932A8 CRC64;

Query Match 5.1%; Score 271.5; DB 1; Length 803;
Best Local Similarity 20.5%; Pred. No. 2.4e-05;
Matches 130; Conservative 63; Mismatches 220; Indels 221; Gaps 22;

364 LEPRGHRKRCFAKVFQSDALQIHLRSHTGPRKCNVCGNRFTTGRGNLKVPHRHRE 423
331 IEKP-----YKCKQCKGKGFSSRRSALNWKVHTAKRYNCECGGAFAFSQASHLDHRLHTG 387
424 KYP-----HYQNNPVRPHLDVYITSSGDPYGNV-----PPE 457
388 EKPFKCDACGKSFRRSHLQSHQVRHTGKPYKCECGKGFICSSNLXIHQVRHTGKPY 447
458 KAEFEATPGGVERKPLVASTALSTSLILTSAGATATAPGLPAFKFVLMKAVER 517
448 KCEE-----CGKGFRRPSSLAHQGVHTGK-SYICTVCGKG-----FTLLSNLQA 492

518 KNRADENTPGSEGSATISGVAESSYATIMQSLKMTSLPWSALLTNHFK-----STGSFP 572
493 HQVHTGKPYKCKECCGKSFRRS-----HYQHLVHTGKPY 530
573 --LPLCARALGASPESSKLOOLVEKIDROGAVAVTSAAGAPTTAPADSSASSGPN 630
531 YKCEICGK--GFSQSSYLIQIHQKHSIEK-----PFK 560
631 CVYCLRLSCPRALRIHNGQNGGERPFCKYCGRAFSTRGNLRAHFVGHKASPARAQN 690
561 CEECGGPFNOSRLQIHQIHNGEKPYKCECGKGFSSRRADLIHCRH-----TGKPYN 616
691 CPICQKFTNAVTLQOHVHMLGQIPNGGTALPEGGAAOENGSEOSTVSAGSFPFOQ 750
617 CEECGKVFRRQASNLAHQRVHSG-----EKPFCECGK-----SFGSA 656
751 SQQPSPEEISEEEDDEEDVTDSDSLAGRSGESGGEKAIYVRGDSSEASGAEEV 810
657 HLQAHQVHTGDKPKYKDE-----CGKGF----- 680
811 GTVAATAATAGKEMDSNETTQSSLPFRPPPSLDQPOPMEGSSGYLGKKEGKPRS 870
681 -----KWSL-----NLDRHQVHTGKPYKCG--ECGKTFSQ 710
871 SSPASALTPEGEATSVTLVEELSLQEAMRKPEGSSSRKACEYCGAFAPSQALEHQKT 930
711 AS-----SLQDHQSVHTGKPYK--CDVCGKVFRRSSQLQSHQRV 748
931 HPKEGPLEFTVCFRQGFLEKATLKKHMLAHNOV 964
749 HTGKRP-YKCEICGKSFRRS-----LTVNHRI 776

RESULT 29
Z084.XENLA
ID Z084.XENLA STANDARD; PRT; 780 AA.
AC P18753;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 01-OCT-1994 (Rel. 30, Last annotation update)
DE Oocyte zinc finger protein XLCOF8.4 (Fragment).
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
OC Xenopodidae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE OF 1-272 FROM N.A.
RX MEDLINE=69345612; PubMed=2503827;
RA Knoechel W., Poeltner A., Koester M., el Baradi T., Niefeld W.,
RA Boumeester T., Pieler T.;
RT "Evolutionary conserved modules associated with zinc fingers in
RT Xenopus laevis.";
RT Proc. Natl. Acad. Sci. U.S.A. 86:6097-6100(1989).
RN [2]
RP SEQUENCE OF 245-780 FROM N.A.
RX MEDLINE=90040698; PubMed=2509712;
RA Niefeld W., el Baradi T., Mentzel H., Pieler T., Koester M.,
RA Poeltner A., Knoechel W.;
RT "Second-order repeats in Xenopus laevis finger proteins.";
RT J. Mol. Biol. 208:639-659(1989).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: M25868; AAF50015.1; -
DR PIR: C33282; C33282.
DR PIR: S06548; S06548.


```
HSSP; P08047; 1SP2.
InterPro; IPR000822; znf_C2H2.
Pfam; PF00096; zf-C2H2; 15.
SMART; SM00355; znf_C2H2; 15.
PROSITE; PS00028; ZINC_FINGER_C2H2_1; 15.
PROSITE; PS50157; ZINC_FINGER_C2H2_2; 15.
Zinc-finger; Metal-binding; DNA-binding; Repeat.
NON_TER 1
T ZN_FING 250 272 C2H2-TYPE.
T ZN_FING 278 300 C2H2-TYPE.
T ZN_FING 306 328 C2H2-TYPE.
T ZN_FING 334 356 C2H2-TYPE.
T ZN_FING 362 384 C2H2-TYPE.
T ZN_FING 390 412 C2H2-TYPE.
T ZN_FING 418 440 C2H2-TYPE.
T ZN_FING 446 468 C2H2-TYPE.
T ZN_FING 474 496 C2H2-TYPE.
T ZN_FING 618 640 C2H2-TYPE.
T ZN_FING 646 668 C2H2-TYPE.
T ZN_FING 674 696 C2H2-TYPE.
T ZN_FING 702 724 C2H2-TYPE.
T ZN_FING 730 752 C2H2-TYPE.
T ZN_FING 758 780 C2H2-TYPE.
T NON_TER 780 780
SQ SEQUENCE 780 AA; 88648 MW; 4880C5BEFC9B3915 CRC64;

Query Match 5.1%; Score 270; DB 1; Length 780;
Best Local Similarity 18.4%; Pred. No. 2,7e-05;
Matches 163; Conservative 81; Mismatches 308; Indels 332; Gaps 30;

149 PLPESTPAAPPP-----PPPPPGVSGHLNP 178
144 PTNPISPKOPPPANGIKEVASCDGROSDCSINPLETOIGTDIPINGYMLIQ 203
179 LILE-----ELRVLOQROIHOMOTEJOICROYLLGSLQGVAPASPSLPGTGTAS 231
204 AIKDVENSPHSRFRNHTHTTE--KQFSHINLHNDLSIHAKKPRPSCGCGCFAG 261
232 STRPL--PLFSPIKPVQTSKTLASSSSSSSSSGAETPKQAFHLYPLGSOHPFSAG 289
262 SSELVHNRHTRVKVF-----SCSOGCKCFSTQTKLHNHTHT--GEKPRSCSE 310
290 VGRSKRPAPSPALPGSIDQLASPHL-----AFPSTTGLLAOLCLGARGL 337
311 CGK-----CPSTPHVARHOKTHGKGPFPSCSE--CGKCF--ARSS 347
338 EATASPGILKPKNGSELGYEVMGLEKPGGRHKRCRCACAVFGSDALQILHRSHTGER 397
348 DVT-----VHRRTHTG-----EKP--YSCSGCGKCFTRSSDLNVRHTHTGK 388
398 PYKCNVCGNRFTTNGMLKYFHHRREKYPHYQMNPHVPEHLDYVTSSGLPYGMSVPE 457
389 PYSCGHGCGFCTTSSSLNVRHTHTGKPY----- 418
458 KAEELAAATPGGVEKRPVASTALSTATESITLTLSTAGTAPGLPAFNKLYLMAVER 517
419 ----- 418
518 KKKADENPPGSGSAISGVAESSTATIMQLSKLMTSLPWNALLTNHPKSTGSFPLPCA 577
419 -----SCSECGKSFPTTSSEFTSHMT----- 439
578 RALGASPSETSKLOOLVEKIDROGAVAVNTSAASGAPTTSAAPSSASSGPNOCVTLRV 637
440 -----HMEK-----PRSCVCGGKC 454
638 LSCPALRLHYGONGGEERPKCKVCGRAPSTRGNILRAHFGVGHKASPARAAQNSCPLCK 656
455 FSKDTHLKXHTHTGKCPSCPCGCGKCFTHNGSLKYLTKHK-----READFCSSKG 506
697 KFTN-AVTLLOOHVRLMLGGQIPNGGTALPRGGGAQENGSEOSTVSGAGSPFOQSQOPS 755
507 NLTNPEISVEHIPP--TNEIKKATISWEGNSQSDYSINSLTEQIQ-----PYTTP 556
```

```
QY 756 PEE-----LSEEEDEE-----EEDVTDEDSLACGSSGGEKALSVRCDSEFASGA 806
D 557 PTMEYNHLIMQDNKYDVYNACHSPLOETFDVT-KHALHKRIDIRRORTQYOLKYDHRNTSD 615
QY 807 EEEVGVAAATAGKEMDSNEKTTQSSLPFRPPRPSIDQOPMGGSS--GVLGKEEG 864
D 616 KP-----LSCSECGKCFSTYHVARHOKTHGKPFSCSECKYARSDLLNVRHTHTG 670
QY 865 GKPERSSPASALTPGEATSVTLVEELSLQEAARKPESSRRACEVCGAAPSQAL 924
D 671 EKPYSCSECGKCFTRSSD-----FNVHRTHTGKRP--YSCSGCGKCFPTSSVL 717
QY 925 EEHQTHPEKGRPLTCVFCRQGFLEKATLKKMLLAHQVQFA 968
D 718 TSHWRHTHTGKRP-FSCTECGKCFSTETLYKYHN-RTHGKERFS 759

RESULT 30
PRDH_HUMAN
ID PRDH_HUMAN STANDARD; PRT; 720 AA.
AC Q9H9D4;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE PR-domain zinc finger protein 17.
GN PRDM17 OR PFM14.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Du Y., Huang S.;
RT "A family of novel PR-domain (PRDM) genes as candidate tumor
RL suppressors.";
RN Submitted (Feb-2001) to the EMBL/GenBank/DBJ databases.
[2]
RN SEQUENCE FROM N.A.
RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Takahashi-Fujii A., Hara H.,
RA Tanase T., Nomura Y., Togiya S., Komai F., Hara R., Takeuchi K.,
RA Arita M., Nabekura T., Ishii S., Kawai Y., Saito K., Yamamoto J.,
RA Wakamatsu A., Nakamura Y., Nagahari K., Masuno Y., Oshima A.;
RT "NDO human cDNA sequencing project.";
RL Submitted (Aug-2000) to the EMBL/GenBank/DBJ databases.
[3]
RN SEQUENCE FROM N.A.
RP TISSUE=Lung, and Uterus;
RC Strausberg R.;
RL Submitted (Oct-2001) to the EMBL/GenBank/DBJ databases.
CC -I- SUBCELLULAR LOCATION: MAY FUNCTION AS A TRANSCRIPTION FACTOR.
CC -I- SUBCELLULAR LOCATION: Nuclear (Potential).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF346626; AAK29075.1; -.
CC EMBL; AK022889; BAB14295.1; -.
CC EMBL; BC013355; AAH13355.1; -.
CC EMBL; BC015708; AAH15708.1; -.
CC HSSP; P08153; 1ZPD.
CC InterPro; IPR000822; znf_C2H2.
CC Pfam; PF00096; zf-C2H2; 10.
CC ProDom; PD000003; znf_C2H2; 1.
CC SMART; SM00355; znf_C2H2; 10.
CC PROSITE; PS00028; ZINC_FINGER_C2H2_1; 10.
CC PROSITE; PS50157; ZINC_FINGER_C2H2_2; 10.
```


Transcription regulation; DNA-binding; Zinc-finger; Metal-binding;
KW Nuclear protein; Repeat.
FT ZN_FING 353 375 C2H2-TYPE.
FT ZN_FING 381 403 C2H2-TYPE.
FT ZN_FING 409 431 C2H2-TYPE.
FT ZN_FING 437 459 C2H2-TYPE.
FT ZN_FING 468 480 C2H2-TYPE.
FT ZN_FING 496 518 C2H2-TYPE.
FT ZN_FING 524 546 C2H2-TYPE.
FT ZN_FING 551 573 C2H2-TYPE.
FT ZN_FING 579 601 C2H2-TYPE.
FT ZN_FING 601 629 C2H2-TYPE.
SQ SEQUENCE 720 AA: 78438 MW: 103CDA7572D17721 CRC64;
Query Match 5.1%; Score 269.5; DB 1; Length 720;
Best Local Similarity 20.6%; Pred. No. 2.6e-05;
Matches 181; Conservative 77; Mismatches 315; Indels 307; Gaps 36;
QY 7 RSSRLG PAG-EPAELGSDASEDHDPQVCAKCAQFTDPT-EFLAHQNA----- 53
Db 19 REPRLG DGLGWNPSGEGCTGGLKDY-----PEPTRILALKSLPRGLALGSLA 68
QY 54 -----CSTDPPVAVYIGQENPNNSASSSEPRRGGHNNPDVMDTEHSNPPDSSGSSVP 105
Db 69 KEORLG TWCYGDDELQGLTWG---PLEEESASKEKEGVK-----PROEENTLS 113
QY 106 TDPYNG EERGESSEGHFLVAATGTAAGG-----GLLASKLGLATPL 150
Db 114 LGP-WGDVACAEQSSGWTSLVQKRLSEGVAVRISERLHLYQVQLVPLGSELLMPQ 172
QY 151 PEEPPAPRRPPRRPPRGVSGHINIRPLLEELRVLQORIQHQMOTQICQVLLLG 210
Db 173 PSESGS -----LTQPGDLKEAVAVTVEBANQOE-----VAS 207
QY 211 LQQTGARA-----SPSELPGTGTAS-----SKPLLPFS-----PI 243
Db 208 PEDDAEPKPCIDPSGQSGIOAENMVSPGLKFPQDRISKDSQPLRLDDGVDEECRA 267
QY 244 KVVQTKTLASSSSSSSSSGAETPKQAFILYHPLGSHRPSAGV---GRSHKPPAP 300
Db 268 Q-AQMPRELQSNATQDDPDS-----GASFSSAKCTOPRHVLAKKLHSP 312
QY 301 SPALPGSTQOLIASPHLAFSTTGLLAAQCLGANGLENTASGELPKNGSGELSYEV 360
Db 313 SQQCP-----PRAKTEP-----GAQSGEPPLTSRSPGRA----- 343
QY 361 MGPLEPRGGHKKCFKAKVGSDSALQILHLSHTGERPKCNCVGNRFTTRGLKVFHR 420
Db 344 -GSSPRQGRYRCGECGKAFLOLCHKHAFVHTGHKRPCTECGKSYSESEFFKAHMLG 402
QY 421 HREKYHVQKNPVP-----EHLDVYITSSGLRYGMSVPRPKAESEATPG 467
Db 403 HR-----GVRRPCCPOCDKAYGTQRDIKEH--QVHSGARPPACD---OCGAFARPPS 451
QY 468 GGVERRPLVASTALSTESTLTLSTAGTAPCLPAFNKEVLMKAVEPKNADEMTTP 527
Db 452 LRHRR-----TIQVPAARA-----PCRPVCGRPL 477
QY 528 GSGGSAISGVAESTATLMQLSKMTSLPSWALLTNHK-STGSPFL--PLCARAL---- 580
Db 478 ANQGS-----LNHMRHLHCEKPEFLCHGCGAFRORG 509
QY 581 -----GASPS-----TSKIQQLVEKIDROGAVVNTSASGAPTTAPAPSSSASS 626
Db 510 NLRGHLRLHTGERYRCRPHCADAPQLPEL--RRLHLSHTGE----- 550
QY 627 GPNQVYICLRLVLSRPLRLHYHGGGERPKVCGRAFSTGNLRAHVGHKASPAAR 686
Db 551 --HLCPVCGKALRPHRLRAHERLHSGERFPCCGGRAYTLATKLRHLKSHLEDKPYR 608
QY 687 AONSCPIQKKFTNAVTLQOHVRHMLGQIPINGGALPEGGGAQENGSDQTVSGAGSF 746
Db 609 ----CPYCGMGY----TLPGSLRRHQLSHRPEAPCSPSPVSAASE-----PRVVLQAE 655

QY 747 PQ-----QSQOPSPPEELSEEEDEEEDVDDEDSLA 782
Db 656 PQLDTHREBEVSPARDVEVTISEQKCFVYPEEPDA 695
RESULT 31
ZFP93_MOUSE
ID ZFP93_MOUSE STANDARD; PRT; 645 AA.
AC Q6116.
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, last sequence update)
DT 16-OCT-2001 (Rel. 40, last annotation update)
DE zinc finger protein 93 (zfp-93).
GN ZFP93.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_Taxid=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CD-1; TISSUE-Testis;
RX MEDLINE=96207307; Pubmed=8617494;
RA Shannon M., Ashworth L.K., Mucenski M.L., Lamerdin J.E., Branscomb E.,
RT Stubs L.;
RT "Comparative analysis of a conserved zinc finger gene cluster on human
chromosome 19q and mouse chromosome 7.";
RL Genomics 33:112-120(1996).
CC -1- FUNCTION: MAY FUNCTION AS A TRANSCRIPTION FACTOR.
CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -1- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-
FINGER PROTEINS.
CC -1- SIMILARITY: CONTAINS 1 KRAB DOMAIN.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb.ch/announce/>
or send an email to license@isb.slb.ch).
CC -----
DR EMBL: U46186; AAB03529.1; -.
DR HSSP: P25490; 1ZNM.
DR MGI: 107611; Zfp93.
DR InterPro: IPR001909; KRAB.
DR InterPro: IPR000822; Znf_C2H2.
DR Pfam: PF00096; zf-C2H2; 13.
DR Pfam: PF01352; KRAB; 1.
DR PRINTS: PR00048; ZINCFINGER.
DR ProDom: PD000003; Znf.C2H2; 11.
DR SMART: SM00349; KRAB; 1.
DR SMART: SM00355; Znf_C2H2; 13.
DR PROSITE: PS50805; KRAB; 1.
DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; 13.
DR PROSITE: PS50157; ZINC_FINGER_C2H2_2; 13.
KW Transcription regulation; DNA-binding; Zinc-finger; Metal-binding;
KW Nuclear protein; Repeat.
FT DOMAIN 8 86 KRAB.
FT DOMAIN 86 86 ZINC_FINGERS.
FT ZN_FING 285 307 C2H2-TYPE.
FT ZN_FING 313 335 C2H2-TYPE.
FT ZN_FING 341 363 C2H2-TYPE.
FT ZN_FING 369 391 C2H2-TYPE.
FT ZN_FING 397 419 C2H2-TYPE.
FT ZN_FING 425 447 C2H2-TYPE.
FT ZN_FING 453 475 C2H2-TYPE.
FT ZN_FING 481 503 C2H2-TYPE.
FT ZN_FING 509 531 C2H2-TYPE.
FT ZN_FING 537 559 C2H2-TYPE.
FT ZN_FING 565 587 C2H2-TYPE.
FT ZN_FING 593 615 C2H2-TYPE.
FT ZN_FING 621 643 C2H2-TYPE.

30	SEQUENCE	645 AA; 73012 MW; 7459CD140F5AF469 CRC64;
	Query Match	5.1%; Score 269; DB 1; Length 645;
	Best Local Similarity	20.1%; Pred. No. 2,5e-05;
	Matches 139; Conservative	60; Mismatches 214; Indels 280; Gaps 23;
1	294 HKPTAPSPALPGSTDOLASPHLAFSPSTGLLAQCL-----GAAGLEA-TPSPGL	345
2	197 HKRDVPS---SGDCDQVI-----FPMT--LTIQHCYREQKAYOCSSRGVEFSDSPSL	245
3	346 LKPKNGSGSLSGEVWG-----PLEPRG-GRHKCRCAKPFSGD	363
4	246 -----ELHQDTLLGKSPVHSTHKDTRHSPSPVLPDPVNHGRKRWCHCGKGFQOS	297
5	384 SALQTLRHSTGEPKPYKCNVCGNFFTRGNLKVHFHREKRYPHOMNPHVPEHLDYVI	443
6	298 SALQTHQKRVHTGKPRPFCDCSCGKGFSSSDINIRRVHTGEPY-----	341
7	444 TSSGLPYGMSVPEKAEKEEAATPGGVERKPLVASTTALSTATESLTLLSTAGTATAPGL	503
8	342 -----KCEVCGKG-----	349
9	504 PAFNKFVLMKAVEPKKADNTPPGSGSALSGVAESSTATLMQSLKMTSLPSMALLTN	563
10	350 --FTQWHLQAHERRHITGEPYKCGDCGKRFS-----CSSMLHTH	387
11	564 HFKSTGSPFLPLCARALGASPSSETSRLQQLVEKIDROGAVAVTSAAGAPTTSAPAPSSS	623
12	388 QRVHTEERKYE-C-----NECGKRFSLSGNLDIHQRY-----H	419
13	624 ASSGNQCVICLRYLSCPRALRLHYGONGGERPRCKYCGRAFTSGRLNRAHFVGHKASP	663
14	420 TGEKPYKCECGCGFSFSSQSHQRVHTGKPRHCYSVCGKMFSTR---SSHFLDHQRIH	475
15	684 AARQNSCPICTCKKFTNNAVTLQOHVHMLGGQIPNGTALPRGGAGAOENGSGQSVVSGA	743
16	476 TGEKPYKCEYCGKRFWMSLSLHSHQSVHTGK-----PKCECGCGFHSAS-----	522
17	744 GSPFOQSOQSPSEELSEEEDEDEEDVDTEDSLGRSGESGKEALISYRGDSEEA	803
18	523 -----SLQAHNSVHTGEPKPCNVQCKQF	546
19	804 SGAESEVGTVAATAATAGKEMDSNEKTTQOSSLPPRRPDSLDQRPOMEGSSGVLGKKEE	863
20	547 S-----KTS-----NLOAHQRVHTGE-----	562
21	864 GKKPRSSSPASALTEGEATSVTLVEELSLQEAARKREPSSSSKACPCVCGAPPSQAA	923
22	563 --KPYKCDTCGKAFS-----OKSSLQVHORIHITGKPRK--CECGKEPRFWSVG	607
23	924 LEHQKTHRKEGPLFTQVCFRCQGFLEKATLTKH	956
24	608 LSSHQRVHTGKPR-YTCCQCGKGFSGASYPFMH	639
RESULT 32		
25	2N45_HUMAN	STANDARD; PRT; 682 AA.
26	NC O02386; P78472;	
27	01-OCT-1993 (Rel. 27, Created)	
28	01-NOV-1997 (Rel. 35, Last sequence update)	
29	16-OCT-2001 (Rel. 40, Last annotation update)	
30	1E Zinc finger protein 45 (BRC1744).	
31	3N ZNF45.	
32	3S Homo sapiens (Human).	
33	3C Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;	
34	3X Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.	
35	3N NCB1;taxID=9606;	
36	3P SEQUENCE FROM N.A.	
37	3A MEDLINE=97220056; PubMed=9067431;	
38	3A Constantinou-Deltas C.D., Bashirides E., Patsalis P.C.,	
39	3A Hadjimarcoiu M., Kriolsei P.M., Ioannou P.A., Roses A.D., Lee J.E.;	

RT	*Complete coding sequence, exon/intron arrangement and chromosome
RT	location of ZNF45, a KRAB-domain-containing gene";
RL	Cytogenet. Cell Genet. 75:230-233(1996).
RM	[2]
RP	SEQUENCE OF 1-400 FROM N.A.
RX	MEDLINE=92217982; PubMed=1559709;
RA	Constantinou-Deltas C.D., Gilbert J., Bartlett R.J., Herbstreith M.,
RA	Roses A.D., Lee J.E.;
RT	"The identification and characterization of KRAB-domain-containing
RT	zinc finger proteins.";
RL	Genomics 12:581-589(1992).
CC	-I- FUNCTION: MAY BE INVOLVED IN TRANSCRIPTIONAL REGULATION.
CC	-I- SUBCELLULAR LOCATION: Nuclear.
CC	-I- SIMILARITY: BELONGS TO THE KRUPEL FAMILY OF C2H2-TYPE ZINC-
CC	FINGER PROTEINS.
CC	-I- SIMILARITY: CONTAINS 1 KRAB DOMAIN.
CC	-----
CC	This SWISS-Prot entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC	the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC	or send an email to license@isb-sib.ch).
CC	-----
DR	EMBL; L75847; AAB05653.1; -
DR	EMBL; M67509; AAA36133.1; -
DR	EMBL; M67509; AAA36134.1; -
DR	HSSP; P08046; IAI1.
DR	TRANSFAC; T04988; -
DR	Genew; HGNC:13111; ZNF45.
DR	MIM; 601781; -
DR	InterPro; IPR001909; KRAB.
DR	InterPro; IPR000822; Znf_C2H2.
DR	Pfam; PF00096; zf-C2H2; 15.
DR	Pfam; PF01352; KRAB; 1.
DR	PRINTS; PR00048; ZINCFINGER.
DR	ProDom; PD000003; Znf_C2H2; 13.
DR	SMART; SM00349; KRAB; 1.
DR	SMART; SM00355; Znf_C2H2; 15.
DR	PROSITE; PS50805; KRAB; 1.
DR	PROSITE; PS00028; ZINC_FINGER_C2H2_1; 15.
DR	PROSITE; PS50157; ZINC_FINGER_C2H2_2; 18.
KW	transcription regulation; DNA-binding; Zinc-finger; Metal-binding;
KW	Nuclear protein; Repeat; Polymorphism.
FT	DOMAIN
FT	8 78
FT	192 662
FT	ZN_FING
FT	192 214
FT	276 298
FT	ZN_FING
FT	304 326
FT	ZN_FING
FT	332 354
FT	ZN_FING
FT	360 382
FT	ZN_FING
FT	388 410
FT	ZN_FING
FT	416 438
FT	ZN_FING
FT	444 466
FT	ZN_FING
FT	472 494
FT	ZN_FING
FT	500 522
FT	ZN_FING
FT	528 550
FT	ZN_FING
FT	556 578
FT	ZN_FING
FT	584 606
FT	ZN_FING
FT	612 634
FT	ZN_FING
FT	640 662
FT	ZN_FING
FT	187
FT	VARIANT
FT	255 255
FT	VARIANT
FT	299 299
FT	VARIANT
FT	303 303
FT	VARIANT
FT	504 504
FT	SEQUENCE
SO	682 AA; 78242 MW; 9C8E536EB00FCBB CRC64;

Query Match 5.1%; Score 269; DB 1; Length 682;
Best Local Similarity 22.9%; Pred. No. 2,6e-05;
Matches 148; Conservative 50; Mismatches 246; Indels 202; Gaps 25;

365 EKPGRHRCRCFAKVEGSDSALQIHLSHTGERPYKCNVCGNRPFTTGNLKVH-FHHR 422
161 EKP---YKGEHCYKSFSSWSSLIQINQRAHAGEKPYKCEKCDNAFRFSSLIQAHQVHSRA 217
423 EKYPHQMNHPREPHDYITS--SGLPYGMSP---PEKAEELAAITPGGVE--RKPL 475
218 KSYTN-----DASYSEFSQSRSLPHHQVPTGENPYKVEECGRNVKSSHCAPL 267
476 VASTTSLAESTLTLSTAGTATAPGLPAFNKFYLMKAVEPKNKADENPPRSEGSAIS 535
268 IVHTG-----EKPYKDE-----C 281
536 GVAESSTATLMQSLKMLSLP-----SW-ALLTNHRK-STGSFPLPLCARALGA 582
282 GVGFSORSYLDVHLKVTGKKPKYKCECGKSFSSMRSLQAHQIHSHTEKPYKCECGKGF 339
583 SPSETKLQ-----QLVEKIDRQ--AVANTV---AASGAPTTAPAPSSSASSG--- 627
340 SFYSYSHLNIHCRITHTGEKPYKCECGKGFVSYSHLQAHQIHSHTEKPYKCECGKGF 399
628 -----PQCVCICLRVLSCPRALRLHYGQHGGERPFKCYGCAFTSGNL 672
400 ASNLIDHQRHTGTEKPYQCQDCAKCGKGFSSDPNHFHVHTGEKPYKCECGKGFSSQ 459
673 RAHFVGHKASPARAQAQNSCPICQKFKETNAVTLQAHVHMLGGQIPNGTALPEGGAQ 732
460 LAHQRHHTGTEKPYK---CGTCGKGFSSSDLNIVHCRHTG-----EKPYK 504
733 NGEOSTVSAGSPQQOOSQPSPEBELSEEEDEEEDVTDSDSLAGRGESEGEK 792
505 CGK-----AFSQESSLIQAHQVHTGEKPYQCAE-----CGKGSVSSQL 543
793 AISRGDSEBASGAEEVGTVAATAATAGKEMDSNEKTQOSSLPPPPPSLDQPPMEQ 852
544 QAHQRCHTGTEKPYKCECGKGFCA----- 568
853 GSSGVUG--GKEBGCKPERSSSPASALTPREGATSVTLVEELSLQEMRKPEGSSSK 910
569 --SNFLAHRGVHTGEKPYKRC-----DVCGKRFORSYLAHQVHTGEKPYK-- 613
911 CEVCGAQPQQAALPEEHOKTHPKRGRLFTVCRCRQGFLEKRLATKKN 956
614 CEBCGKVFSSWSSYLAHQVHTGEKPYKCECGKGFSSWSSSLI 658

RESULT 33
2184_HUMAN STANDARD: PRT; 751 AA.
AC Q99676; O60792;
DT 15-JUL-1998 (Rel. 36, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
GN ZNF184.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Phillips S.;
RN Submitted (MAY-1998) to the EMBL/Genbank/DBJ databases.
RP SEQUENCE OF 26-751 FROM N.A.
RC TISSUE=Placenta;
RX MEDLINE=97250463; PubMed=9073517;
RA Goldwurm S., Menzies M.L., Banyer J.L., Powell B.L.W.,
RA Jazwinska E.C.;

RT *Identification of a novel Krueppel-related zinc finger gene (ZNF184)
RT mapping to 6p21.3".
RL Genomics 40:486-489(1997).
CC - FUNCTION: MAY BE INVOLVED IN TRANSCRIPTIONAL REGULATION.
CC - SUBCELLULAR LOCATION: Nuclear (Probable).
CC - TISSUE SPECIFICITY: PREDOMINANT EXPRESSION IN TESTIS.
CC - SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-FINGER PROTEINS.
CC - SIMILARITY: CONTAINS 1 KRAB DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AL021918; CAA17278.1; -.
DR EMBL; U66561; AAC51180.1; -.
DR HSSP; P08046; 1A1H.
DR Genew; HGNC:12975; ZNF184.
DR MIM; 602277; -.
DR InterPro; IPR001909; KRAB.
DR InterPro; IPR000822; Znf_C2H2.
DR Pfam; PF00096; zf_C2H2; 19.
DR Pfam; PF01352; KRAB; 1.
DR PRINTS; PRO0048; ZINCFINGER.
DR PRODom; PD000003; Znf_C2H2; 18.
DR SMART; SM00349; KRAB; 1.
DR SMART; SM00355; Znf_C2H2; 19.
DR PROSITE; PSS0805; KRAB; 1.
DR PROSITE; PSS00028; ZINC_FINGER_C2H2_1; 19.
DR PROSITE; PSS0157; ZINC_FINGER_C2H2_2; 19.
KW Transcription regulation; DNA-binding; Zinc-finger; Metal-binding;
KW Nuclear protein; Repeat.
KW DOMAIN 28 99
FT FT 222 748 KRAB. ZINC FINGERS.
FT ZN_FING 222 244 C2H2-TYPE.
FT ZN_FING 250 272 C2H2-TYPE.
FT ZN_FING 278 300 C2H2-TYPE.
FT ZN_FING 306 328 C2H2-TYPE.
FT ZN_FING 334 356 C2H2-TYPE.
FT ZN_FING 362 384 C2H2-TYPE.
FT ZN_FING 390 412 C2H2-TYPE.
FT ZN_FING 418 440 C2H2-TYPE.
FT ZN_FING 446 468 C2H2-TYPE.
FT ZN_FING 474 496 C2H2-TYPE.
FT ZN_FING 502 524 C2H2-TYPE.
FT ZN_FING 530 552 C2H2-TYPE.
FT ZN_FING 558 580 C2H2-TYPE.
FT ZN_FING 586 608 C2H2-TYPE.
FT ZN_FING 614 636 C2H2-TYPE.
FT ZN_FING 642 664 C2H2-TYPE.
FT ZN_FING 670 692 C2H2-TYPE.
FT ZN_FING 698 720 C2H2-TYPE.
FT ZN_FING 726 748 C2H2-TYPE.
FT ZN_FING 727 27 A->S (IN REF. 2).
FT CONFLICT 295 295 Q->R (IN REF. 2).
FT CONFLICT 359 359 E->G (IN REF. 2).
FT CONFLICT 379 381 OHQ->P (IN REF. 2).
FT CONFLICT 498 498 R->G (IN REF. 2).
FT CONFLICT 526 526 Q->G (IN REF. 2).
FT CONFLICT 533 533 K->E (IN REF. 2).
FT CONFLICT 547 549 KHE->Q (IN REF. 2).
FT CONFLICT 549 549 R->K (IN REF. 2).
FT CONFLICT 746 746 R->K (IN REF. 2).
SQ SEQUENCE 751 AA; 86174 MW; 249DFB302C6BC97D CRC64;

Query Match 5.1%; Score 268; DB 1; Length 751;
Best Local Similarity 21.2%; Pred. No. 3,2e-05;
Matches 136; Conservative 43; Mismatches 220; Indels 242; Gaps 21;
365 EKPGRHRCRCFAKVEGSDSALQIHLSHTGERPYKCNVCGNRPFTTGNLKVH-FHHR 424


```
Db 303 EKP---YKCECCGAFSQRTHLYVQHRIHTGKRPYCNCEGKAFFSQRG-----HFMHQK- 354
2Y YPVQWMPHPEVPEHLIDYVITSSGLPYGMSVPEPKAEFEAAATPGGVRKRLVASTTALSA 484
Db 355 -HTGKPEFCDE-CDKTFYTS-----THLTQ 379
2Y 485 TESL-----TLSTAGATATAPGLPAFNKFLMKAVEPKNAKADENTPPGSEGAISGVAE 539
Db 380 HOKIHGKETYKKNCEGKA-FNGPSTFIRHHNIHTGKPEYECNE-----CGKAF 427
2Y 540 SSTATLMQLSKLMT-----SLPVALLTNNHFK-STGSFPLPL--CARAL----- 580
Db 428 SOHSNLTQHQKHTGKRPYDCAECGKSFYSWSSLAQHLKIHGKRPYKKNCEGKAFFSYCS 487
2Y 581 -----GASPEYSKILQOLVEKIDROGAVAVYSAAGAPTSAAP 620
Db 488 SLTQHRHRIHTREKPEFSECGKAFSYLSNLNOHQKHTQEKAECEGKAFFIRSSSLAK 547
2Y 621 SSSASSG--PNOCVICLRVLSCPRLRLHYGOHGERPRPKCKVCGRAFYTRGNLRAHFVG 678
Db 548 HERIHGKPEYQCHEGCKTFYSYSSLIOHRKIHTGERPYKKNCEGKAFF-----NQHILTQ 603
2Y 679 HKASPARAQNSCPICQKKFTNNATYLOQHVHMLGQITNGGIALPEGGGAQDNGSEQS 738
Db 604 HKRIHGAKPYEACGKAFHRCSSLAQHQKH----- 636
2Y 739 TVSGAGSFPOQSQPPEELSEEEDEEEDVDDEDLAGRSGESGKEAISVRG 798
Db 637 -----TEKKP----- 641
2Y 799 DSEBAGAEVEVGTAAATAAGKEMDSNEKTTQSSLPPEPPPSLDQPMQSGSSGVL 858
Db 642 -----YQCNKCEKTFSSQS-----HLTQHQRIHTGE---- 667
2Y 859 GGEKEGKPRSSSPASALTPPEGATSVTLVEBLSIQEMRKKEPSSSSKKAECVCCQAF 918
Db 668 -----KPYKCECDKRAFS-----RSTHLTEHQNTHTGKRPYN--CNECKRKF 707
2Y 919 PQOALPEHQKTHPKEGPLFTVCPRQGFLEKATLKKHML 959
Db 708 SQSTYLIQHQRIHSGKRP-FGCNDCGKSFRRSALNKHQRL 747
RESULT 34
ID Z071_XENLA STANDARD; PRT; 898 AA.
AC P18751;
2T 01-NOV-1990 (Rel. 16, Last Created)
2T 01-NOV-1990 (Rel. 16, Last sequence update)
2T 01-OCT-1994 (Rel. 30, Last annotation update)
2E Oocyte zinc finger protein XLCOF7.1 (Fragment).
2C Xenopus laevis (African clawed frog).
2C Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
2C Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
2C Xenopodinae; Xenopus.
2X NCBI_TaxID=355;
2X [1]
2X SEQUENCE OF 1-311 FROM N.A.
2X MEDLINE=89345612; PubMed=2503827;
2X Knoechel W., Poeling A., Koester M., el Baradi T., Niefeld W.,
2X Bouwmeester T., Pieler T.,
2X "Evolutionary conserved modules associated with zinc fingers in
2X Xenopus laevis.";
2X Proc. Natl. Acad. Sci. U.S.A. 86:6097-6100(1989).
2X [2]
2X SEQUENCE OF 284-898 FROM N.A.
2X MEDLINE=90040698; PubMed=2509712;
2X Niefeld W., el Baradi T., Wentzel H., Pieler T., Koester M.,
2X Poeling A., Knoechel W.;
2X "Second-order repeats in Xenopus laevis finger proteins.";
2X J. Mol. Biol. 208:639-659(1989).
2X -----
```

```
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M25866; AAA50013.1; -.
DR PIR; A33282; A33282.
DR PIR; S06546; S06546.
DR HSSP; P08047; ISP2.
DR InterPro; IPR000822; Znf_C2H2.
DR Pfam; PF00096; zf-C2H2; 21.
DR SMART; SM00355; znf.C2H2; 21.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 21.
DR PROSITE; PS0157; ZINC_FINGER_C2H2_2; 21.
KW Zinc-finger; Metal-binding; DNA-binding; Repeat.
FT ZN_FING 289 311
FT ZN_FING 342 364 C2H2-TYPE.
FT ZN_FING 370 392 C2H2-TYPE.
FT ZN_FING 398 420 C2H2-TYPE.
FT ZN_FING 425 448 C2H2-TYPE.
FT ZN_FING 426 448 C2H2-TYPE.
FT ZN_FING 454 476 C2H2-TYPE.
FT ZN_FING 481 504 C2H2-TYPE.
FT ZN_FING 482 504 C2H2-TYPE.
FT ZN_FING 510 532 C2H2-TYPE.
FT ZN_FING 537 561 C2H2-TYPE.
FT ZN_FING 538 561 C2H2-TYPE.
FT ZN_FING 567 589 C2H2-TYPE.
FT ZN_FING 595 617 C2H2-TYPE.
FT ZN_FING 623 645 C2H2-TYPE.
FT ZN_FING 651 673 C2H2-TYPE.
FT ZN_FING 679 701 C2H2-TYPE.
FT ZN_FING 707 729 C2H2-TYPE.
FT ZN_FING 735 758 C2H2-TYPE.
FT ZN_FING 764 786 C2H2-TYPE.
FT ZN_FING 792 814 C2H2-TYPE.
FT ZN_FING 820 840 C2H2-TYPE.
FT ZN_FING 848 870 C2H2-TYPE.
FT ZN_FING 876 898 C2H2-TYPE.
FT NON_TER 898 898
SQ SEQUENCE 898 AA; 101167 MW; 795D806B5696B0B6 CRC64;
Query Match 5.1%; Score 268; DB 1; Length 898;
Best Local Similarity 19.3%; Pred. No. 3.8e-05;
Matches 201; Conservative 91; Mismatches 371; Indels 378; Gaps 40;
QY 17 EPALGGDASEEDHPQVCAKCAQFTDPTFLAHQNA---CSTDPPVMYIIGQENPNNS 73
Db 126 EPQQL-----RPQC-----EYKDESNVTAHMEATLCCNSD-----GNFINENP 165
QY 74 --SASSPRREGHNPNQVMDIENHSPPDSGSSVPDPPTWGPERRGESSGHFLVAATGTA 131
Db 166 EISPGEQPPANGIKKEATSSSEGNQSDCSINPLIEEIOGTD----- 207
QY 132 AGGGGGLILASPKLGATPLPESTPAPPPPPPPPGVSGHLNIPL-----ILLEE 183
Db 208 -----ITPT-----IMGSLNLSLDNYSIDEKKEE 233
QY 184 LRYLQQRQIHOMQNTQICQVLLLSIGQTVGAPASPELPGTASTKPLPLPESP1 243
Db 234 ETSQGRNQSDCSIT-----PLTEQIQGTDTPTPIGSCSLKDNKYDGNPHMSPK 282
QY 244 KPYQT-----SKTLASSSSSSSSSGAETPQQAFFHLX-----HPL 279
Db 283 NTLRRRYSCNECHEYLLHKRDPFGKHOMTHKREKSFSCSCG-----KCPILNQLCPDRHQY 337
QY 280 GSOHPSAGVGGRSHKPTPAVSPALPGSTDQLIASPHLAFPST-----TGLLAQCL 331
Db 338 GK-KPSCSKGKC-----FAFLSDLTVYHRIHTGERPPSCS 373
```


332 GAARGLLEATASPELL-----KPKNGSGELSYGEV-----MCPLEKPGGR 370
 Db ECGKEF---TRPALLIHHHTHTGKEKPFSCSECKGKSSLVHQRTHIG--EKP--- 425
 371 HKRCRCVKVGSQSLQILHRSHTGERPYKNCVCGNRFTTGNLKVHFRHREKRYPHVM 430
 426 FCCSECDKCFASSELEINHQRTHGKEKPFSCSECKGKFTNHS-----HFAHHQ----- 473
 431 NHPREHIDYITSSGLPRGMSVPRPEKAEEDATPGGVVEKPLVASTALATESLT 490
 474 -----MHHG-----EKPCCSKCGKCFASSSDL 497
 491 ISTSAGTAPGLPAEFKFLMKAVEBKNADENTPGSE--GSAISGVESSTATLMQ-- 547
 498 -----TFHR-----RTHREKTFSCSECKGKFSNHLARHQMHTG 534
 548 -----LSKMTSLPSWALTLNHFSTGSFPLPCALGASPESTKL--QOLVERIDRQ 600
 535 EKPFCCSECKGKFSSSGLTAHQORTMKVKPFSCSACGKFSNRSHLIRHQMHTGKEK- 593
 601 GAVAVTSAASGATTSAPARSSASSGPNOCVYICLRVLSCPALRLHYGONGGERPKCK 660
 534 -----PFCSECKKCFSPNSNLAHQMTHTGKEKPFSCS 626
 661 VCGRASTRGNLRAHFYGHKASPARAQNCGPICOKKFTNAVTLQOHVRMLGOLPNGG 720
 627 ECGKCFASSSDLTFHHHTH---TGKRFSCSECKGKCYKSKSLVHQRHTH--GEKPFSC 681
 721 TALPEGGAQENGSEGSTVSGAGSPQOQSOQSPSEBELSEEEDEEEDVYDEDS 780
 682 SKCDKCFASSSELEINHQRTHGKEKAF-----SCSECKGKCFTRSQ 721
 781 LAGRGSESGEKALISVAGDSEAGAEVEVTAATAATGKEMDSNEKTKOQSSLP PPP 840
 722 LSRHQMHTGKEKISC--PECEECF-----VSSQLTAHQQAHRMVK 762
 841 PSLDPOQPMEOGSSVGLGKEGK--PERSSPASALTPEGEATSVTLVEELSLQEAM 898
 763 PRSC-----LECGKFSNRSNFRHQMHTG-----KPFSCDECR 798
 899 RKEPGESSSRK-----ACEVCGQAPPSQALAEHQKTHPEKGLFTVCFCROGF 947
 799 KGRSNOSSLARHQMHTGKEKPFSCSECKGKFSNQSLARHQMHTGKEK--FSCSECKAGF 857
 948 LERATLKKHMLLAHHOVRPA 968
 858 SNOGLARHQ-MTHGKEKPPA 877

RESULT 35
 2192_HUMAN STANDARD: PRT; 578 AA.
 AC Q15776; Q9H4T1;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Zinc finger protein 192 (ZNF192).
 GN ZNF192.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 OX NCBI_TaxID=9606;
 RX [1]
 RP SEQUENCE FROM N.A. AND VARIANT LEU-163.
 RC TISSUE=Ovary;
 ME MEDLINE=97386587; PubMed=9244436;
 RA Lee P.L., Gelbart T., West C., Adams M., Blackstone R., Beutler E.;
 RT "Three genes encoding zinc finger proteins on human chromosome 6p21.3:
 RT members of a new subclass of the Kruppel gene family containing the
 RT conserved SCAN box domain.";
 RL Genomics 43:191-201(1997).
 RN [2]
 RP SEQUENCE FROM N.A.

RA Williams S.;
 RL Submitted (SEP-2001) to the EMBL/Genbank/DBJ databases.
 RN [3]
 RP PARTIAL SEQUENCE FROM N.A.
 RC TISSUE=Ovary;
 RX MEDLINE=96230927; PubMed=8673473;
 RA Beutler E., Gelbart T., West C., Kuhl W., Lee P.;
 RT "A strategy for cloning the hereditary hemochromatosis gene,";
 RT Blood Cells Mol. Dis. 21:206-216(1995).
 CC - FUNCTION: MAY BE INVOLVED IN TRANSCRIPTIONAL REGULATION.
 CC - SUBCELLULAR LOCATION: Nuclear (potential).
 CC - SIMILARITY: BELONGS TO THE KRUPPEL FAMILY OF C2H2-TYPE ZINC-
 CC FINGER PROTEINS.
 CC - SIMILARITY: CONTAINS 1 SCAN BOX.
 CC - SIMILARITY: CONTAINS 1 SCAN BOX.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: U57796; AAB02260.1; -;
 DR EMBL: AL358933; GAC15901.1; -;
 DR EMBL: U88080; AAC51656.1; -;
 DR EMBL: U88079; AAC51656.1; JOINED.
 DR HSSP: P07248; 1PAA.
 DR Genew: HGNC:12983; ZNF192.
 DR MIM: 602240; -;
 DR InterPro: IPR001909; KRAB.
 DR InterPro: IPR003309; Treg_SCAN.
 DR InterPro: IPR000822; Znf_C2H2.
 DR Pfam: PF00096; Zf-C2H2; 9.
 DR Pfam: PF01352; KRAB; 1.
 DR Pfam: PF02023; SCAN; 1.
 DR PRINTS: PR00048; ZINCFINGER.
 DR ProDom: PD000003; Znf_C2H2; 8.
 DR SMART: SM00349; KRAB; 1.
 DR SMART: SM00431; LER; 1.
 DR SMART: SM00355; Znf_C2H2; 9.
 DR PROSITE: PS50805; KRAB; 1.
 DR PROSITE: PS50804; SCAN_BOX; 1.
 DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; 9.
 DR PROSITE: PS50157; ZINC_FINGER_C2H2_2; 9.
 KW Transcription regulation; Zinc-finger; Metal-binding; DNA-binding;
 KW Nuclear protein; Repeat; Polymorphism.
 FT DOMAIN 51 133
 FT DOMAIN 220 316
 FT DOMAIN 322 568
 FT ZN_FING 322 344
 FT ZN_FING 350 372
 FT ZN_FING 378 400
 FT ZN_FING 406 428
 FT ZN_FING 434 456
 FT ZN_FING 462 484
 FT ZN_FING 490 512
 FT ZN_FING 518 540
 FT ZN_FING 546 568
 FT VARIANT 163 163
 FT
 FT CONFLICT 34 34 A -> D (IN REF. 2).
 FT CONFLICT 113 113 D -> E (IN REF. 2).
 FT
 FT SEQUENCE 578 AA; 65757 MW; 5914F5439F4B8AB8 CRC64;
 SQ
 Query Match 5.1%; Score 267; DB 1; Length 578;
 Best Local Similarity 22.7%; Pred. No. 2.7e-05;
 Matches 137; Conservative 61; Mismatches 207; Indels 198; Gaps 23;
 Oy 177 IPILELRVLQQRQIHOQMTEQIC-----ROYULLS-----LQGTQAP 218
 Db 100 LTIPEELQTLVKD--HOLENCEEVVTLLEDLERQIDILGRPVASRVHGRVLEEVVHS 157

yy	219	AISSTLPGT---	GTA	SS	TPLP	LFP	SP	IKVQ	TSKTLA	SSSSSSS-----	SSGA	266
yj	158	ASAPPPNTOLOSEA	VOHKS	PV	OES	ERM	S	TOS	PSR	SKSGSSGDD	EMATLL	217
yl	267	T-----	PKOA-	FHYL	HP	LGSO	HFFS	AGVGRSHK	PTPAP	SFALPG		306
yj	218	TLEKTEDMAVSLI	RREMLD	PSQKD	C	DRNP	REPM	FRNMFS	LSLGG	EFTSR	NRELA-----	270
zy	307	STDOLIAS---	PHLAF	PTTG	LGLAAOCLG-	AARCLE	ATASFG	LKPKN	SGELSTGEYWG			362
yj	271	-SKOVISTGIQPH-	GETA	KACNGDVIR	GLEHEARDLL-							307
zy	363	PLEKGCG-----	RHKCR	PAKYKFGS	DSALOIH	RSHTGER	PKYNVC	NGNRPTTG	NLKAV			416
Db	308	RLEROGNPTQERR	HKDCDECKSF	AOSSGLYRRMR	IHTGEPYC	GCAGKAFSYNSALS						367
zy	417	H--FHHRREKYPHV	OMNPNRPVPEHL	DVITTSGLPYGMS	VDPKEKAEEBA	ATPDGGVERKP						474
Db	368	HODHNKKVRV-HCK				ECGKA	FSQNT-----					398
zy	475	LVASTTALSATES	LTLLSTS	SAGTATA	AGLPANFKNFVLMKAVE	PPKAKADENTPRGS	SGSAI					534
Db	399	IHTGEPYOCNOCK	KA	FSOSAG-----	LLHQRIHSGE	RYECNE-----						438
oy	535	SGVAESTTATMOL	SKIMTSLP	SMALLTNHFK	STGPSPLPICARALGASP	SETSKLOOLY						594
Db	439	CGKA	FSHSHLIGHQRHT									471
zy	595	EK----	IDROGA	VAVTSA	ASGAPTTSA	PABSSASSGPNOCVICTLRVLSC	PRALRLHYGO					650
Db	472	RSSH	LIGHQ-----			RSHTG	EKRYPKCCECGRASFQSKGLIEHQRI					511
zy	651	HGGERPFKCCVCG	RATSTRGNL	LANHFVGHKAS	PARAQNSCPICQKKT	FNATYLQHYVM						710
Db	512	HNGERYKCKECKE	CAKANGNTGLIQH	LRHIIH----	TGEPYCNEC	CAFLQRSSLI	NHQRI					567
oy	711	HLG	713									
Db	568	HSG	570									

RESULT 36
 zfp95_HUMAN
 ID zfp95_HUMAN STRAND: PRT; 839 AA.
 AC OGYZL8:
 DT 16-OCT-2001 (rel. 40, Created)
 DT 16-OCT-2001 (rel. 40, last sequence update)
 DT 15-JUN-2002 (rel. 41, last annotation update)
 DE zinc finger protein zfp-95.
 GN zfp95 OR KIA1015.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 NCBI_TaxId=9606;
 [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20054365; PubMed=10585779;
 RA Dreyer S.D., Zheng Q., Zabel B., Witterpacht A., Lee B.,
 "Isolation, characterization, and mapping of a zinc finger gene,
 zfp95, containing both a SCAN box and an alternatively spliced KRAV A
 domain.";
 RL Genomics 62:119-122(1999).
 [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RA MEDLINE=99246063; PubMed=10231032;
 RA Nagase T., Ishikawa K.-I., Suyama M., Kikuno R., Hirosewa M.,
 Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
 RT "Prediction of the coding sequences of unidentified human genes.
 XIII. The complete sequences of 100 new cDNA clones from brain which
 code for large proteins in vitro".
 RL DNA Res. 6:63-70(1999).

```
CC -!- FUNCTION: MAY BE INVOLVED IN TRANSCRIPTIONAL REGULATION.
CC -!- SUBCELLULAR LOCATION: Nuclear (potential).
CC -!- SIMILARITY: BELONGS TO THE KRUEPEL FAMILY OF C2H2-TYPE ZINC-
CC FINGER PROTEINS.
CC -!- SIMILARITY: CONTAINS 1 KRAB DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 SCAN BOX.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL: AF170025; AAF2A219.1; -.
DR EMBL: AB023232; BAA76859.1; -.
DR HSSP: P07248; ZADR.
DR TRANSFAC: T04978; -.
DR Genew: HGNC:12867; ZFP95.
DR InterPro: IPR001909; KRAB.
DR InterPro: IPR003309; Treg_SCAN.
DR Interpro: IPR000822; ZnF_C2H2.
DR Pfam: PF00096; zf-C2H2; 12.
DR Pfam: PF01352; KRAB; 1.
DR Pfam: PF02023; SCAN; 1.
DR Prodom: PD000003; ZnF_C2H2; 9.
DR SMART: SM00349; KRAB; 1.
DR SMART: SM00431; LER; 1.
DR SMART: SM00355; ZnF_C2H2; 12.
DR PROSITE: PSS0805; KRAB; 1.
DR PROSITE: PSS0804; SCAN_BOX; 1.
DR PROSITE: PSS0082; ZINC_FINGER_C2H2_1; 12.
DR PROSITE: PSS0157; ZINC_FINGER_C2H2_2; 13.
KW Transcription regulation; DNA-binding; Zinc-finger; Metal binding;
KW Nuclear protein; Repeat.
FT DOMAIN 51 132 SCAN_BOX.
FT FT 218 291 KRAB.
FT FT 346 823 ZINC_FINGERS.
FT FT 346 368 C2H2-TYPE.
FT FT 374 396 C2H2-TYPE.
FT FT 402 424 C2H2-TYPE.
FT FT 430 452 C2H2-TYPE.
FT FT 449 571 C2H2-TYPE.
FT FT 577 599 C2H2-TYPE.
FT FT 603 627 C2H2-TYPE.
FT FT 633 655 C2H2-TYPE.
FT FT 661 683 C2H2-TYPE.
FT FT 717 739 C2H2-TYPE.
FT FT 745 763 C2H2-TYPE (ATYPICAL).
FT FT 773 795 C2H2-TYPE.
FT FT 801 823 C2H2-TYPE.
SQ SEQUENCE 839 AA; 96903 MW; F2E082CF6610806D CRC64;

Query Match 5.0%; Score 266; DB 1; Length 839;
Best Local Similarity 19.9%; Pred. No. 4.3e-05;
Matches 121; Conservative 76; Mismatches 199; Indels 212; Gaps 22;
```


Db	586	TEKPIIMCDKCGSFKKRTYFKNHLHLTHIOAVANRRKRCFCEVCECKKALLNHQLSHS	645
QY	654	ERPEKCVKVGAFSTRGNLRAHF-VGKH-ASPAARAONSCPTQCKRETNATVLOOHVHM	711
Db	646	DKPFCSCFPYRTFRFDLLSHVAVKHTGAKPPA-----CEYCHFSTRHKNLRLHVR	700
QY	712	LGQQLPNCGTALPEGGAQOENQSEDS---TVSGASSFPQOQSOQSP-----EELISE	762
Db	701	HASSFEEMGRHPE-----EPPSRRRPFFSLQOIELKQHSMAAPPSPSSPGPEPI	754
QY	763	EEEDDEDEEDVDTEDEDSLAGE-----RSESCEGKAISVRGDEEASGAEEVGTAA	816
Db	755	EATTTQSSAPFLSCSDTLGGATIIYQCGAEESTMATQVTLDDLMLMSQORELGITALQ	814
QY	817	ATAGKEMD-----SNEKTQQ-----SSLPPPPPPD-----	842
Db	815	VAVVSEDEVEAGLASPGGQPSPEGATPQVYVTLHVAERPGGAAMESQCGPDLPIITLAP	874
QY	843	-----SLDQOQPMQGGSS--GVLGKREBGKPEBSSSPASALTPGCE---ATSTVLY	889
Db	875	PFGGTGYSTVITAPMEEGTSAPCTPYSEEPAGEAAQVAVVSDTLKEAGTHYIMATDGTOL	934
QY	890	EELSL-----OEMAKREP	902
Db	935	HHIELTLADGSIFFPSPDALASGAKWPLQCGGLPRDGPPEPPSPAKTTCVGDSSQSSASSP	994
QY	903	GES-----SSRKACEVCGAAPPQGALEEHQKTHPKEGPLTFCVC-	943
Db	995	ATSKALGLAVPPSPSAATVAASKKFKCKICALAPFGAEEMSHKRAHAGP-ARKCPDCP	1053
QY	944	-----RQGFLEKATILKKHM-----LLAHQVQPARPH	970
Db	1054	FSARQPEVRAHMAQHSSLRPHQCSQCSFASKNKKDLRRMLTHTTKKPEFACH	1106
RESULT 39			
1D	2337_HUMAN	STANDARD:	PRT; 751 AA.
AC	Q9Y3M9; Q9Y3Y5;		
DT	15-JUN-2002 (Rel. 41, Created)		
DT	15-JUN-2002 (Rel. 41, Last sequence update)		
DT	15-JUN-2002 (Rel. 41, Last annotation update)		
DE	zinc finger protein 337.		
GN	ZNF337.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	SEQUENCE FROM N.A.		
RP	MEDLINE=21638749; PubMed=11780052;		
RX	Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R.,		
RA	Jones M., Stavrides G., Almeida J.P., Babage A.K., Baguley C.L.,		
RA	Bailey J., Barlow K.F., Bates K.N., Beard L.M., Beare D.M.,		
RA	Beasley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J.,		
RA	Buck D., Burrill W.D., Butler A.P., Carder C., Carter N.P.,		
RA	Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M.,		
RA	Clegg S., Cobley V.E., Collier R.E., Connor R.E., Corby N.R.,		
RA	Coulson A., Coville G.J., Deadman R., Dhami P.D., Dunn M.,		
RA	Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,		
RA	Graham D.V., Griffiths C., Griffiths M.N.D., Gwilliam R., Hall R.E.,		
RA	Hammond D., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J.,		
RA	Huckle E., Hunt A.R., Hunt S.E., Jakosch K., Johnson C.M., Johnson D.,		
RA	Key M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,		
RA	Lehvasalo M.H., Leverhna M.A., Lloyd C., Lloyd D.M., Lovell J.D.,		
RA	Marsh V.L., Martin S.L., McConachie L.J., McElay K., McKurray A.A.,		
RA	Milne S.A., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T.,		
RA	Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I.,		
RA	Phillimore B.J.C.T., Patelalingam S.R., Plumb R.W., Ramsay H.,		
RA	Rice C.M., Ross M.T., Scott C.E., Sehra H.K., Showkneen R., Sims S.,		
RA	Skuce C.D., Smith M.L., Soderlund C., Steward C.A., Sulston J.E.,		
RA	Swann R.M., Sycamore N., Taylor R., Tee L., Thomas D.W., Thorpe A.,		

```

RA Tracey A., Tromans A.C., Vaudin M., Walli M., Wallis J.M., Williams S.A.,
RA Whitehead S.L., Whitaker P., Willey D.L., Williams L., Williams S.A.,
RA Wilming L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,
RA Rogers J.;
RT "The DNA sequence and comparative analysis of human chromosome 20.";
RL Nature 414:865-871(2001).
RP
RP SEQUENCE OF 256-751 FROM N.A.
RC TISSUE-BRAIN;
RA Wandurt R., Heubner D., Mewes H.-W., Gassenhuber J., Wiemann S.;
RL Submitted (MAY-1999) to the EMBL/Genbank/DBJ databases.
CC -1- FUNCTION: MAY FUNCTION AS A TRANSCRIPTION FACTOR.
CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -1- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-
CC FINGER PROTEINS.
CC -1- SIMILARITY: CONTAINS 1 KRAB DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL, AL031673; CAB43548.2; -.
DR EMBL, AL049942; CAB43216.1; -.
DR HSSP; P08048; 7ZNF.
DR Genew; HGNC:15809; ZNF337.
DR InterPro; IPR001909; KRAB.
DR InterPro; IPR000822; Znf_C2H2.
DR Pfam; PF00096; Zf-C2H2; 37.
DR Pfam; PF01352; KRAB; 1.
DR ProDom; PD000003; Znf_C2H2; 4.
DR SMART; SM00349; KRAB; 1.
DR SMART; SM00355; ZNF_C2H2; 19.
DR PROSITE; PS00805; KRAB; 1.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 19.
DR PROSITE; PS00157; ZINC_FINGER_C2H2_2; 20.
KW Transcription regulation; Zinc-finger; Metal-binding; Nuclear protein;
KW DNA-binding; Repeat.
FT DOMAIN 12 83 KRAB.
FT FT 180 735 ZINC_FINGERS.
FT FT 180 202 C2H2-TYPE (DEGENERATE).
FT FT ZN_FING 208 230 C2H2-TYPE.
FT FT ZN_FING 236 258 C2H2-TYPE.
FT FT ZN_FING 264 286 C2H2-TYPE.
FT FT ZN_FING 292 314 C2H2-TYPE.
FT FT ZN_FING 320 342 C2H2-TYPE.
FT FT ZN_FING 348 370 C2H2-TYPE.
FT FT ZN_FING 376 398 C2H2-TYPE.
FT FT ZN_FING 404 426 C2H2-TYPE.
FT FT ZN_FING 432 454 C2H2-TYPE.
FT FT ZN_FING 460 482 C2H2-TYPE.
FT FT ZN_FING 488 510 C2H2-TYPE.
FT FT ZN_FING 516 538 C2H2-TYPE.
FT FT ZN_FING 544 566 C2H2-TYPE.
FT FT ZN_FING 572 594 C2H2-TYPE.
FT FT ZN_FING 600 622 C2H2-TYPE.
FT FT ZN_FING 628 650 C2H2-TYPE.
FT FT ZN_FING 656 679 C2H2-TYPE.
FT FT ZN_FING 685 707 C2H2-TYPE.
FT FT ZN_FING 713 735 C2H2-TYPE.
FT FT CONFLICT 476 476 E -> A (IN REF. 2).
FT FT CONFLICT 653 653 E -> V (IN REF. 2).
SQ SEQUENCE 751 AA; 86874 MW; 482D9F9A57EC980E CRC64;
Query Match 5.0%; Score 265; DB 1; Length 751;
Best Local Similarity 19.0%; Pred. No. 4.3e-05;
Matches 167; Conservative 92; Mismatches 269; Indels 352; Gaps 28;
QY 137 GLIASPKL-----GATPLRPPESTRPAPRRPPPPPPVGGSHLNIPILBELRVQQR 190
|:::|::| | | | | | | | | | | | | | | | | | | | | | | | | | | |

```


2b	52	GILHSKPELIRLEQGEVPMGEERRR-----PGPCAGIYAENVLR--KNILGIAHOR	102
2Y	191	QIHOMOMTEOICROVLLIGSGTVGAPASPELBTGTGASSTKPLLPSPKIPVOTSK	250
2b	103	Q-QQLOPDSQFS-----DTAGQEKESTKPAFSSPPLAHAVSSR	144
2Y	251	TLAASSSSSSSSSGAETPKQAFPHLYHPLGSOHPRSSAGVGRSHKPTPAPSPALPGSTDQ	310
2b	145	RNSVYEIESQOGRENPT-----IDK	167
2Y	311	LASPLALPPTSTGILAAOCLGAARGLFATASPGILKPKNSGSELSTYGEVMPLEKGGCR	370
2b	168	VL-----KGIE-NSRMGAFKCAERQODFRKMMVLIHKKAHSR	204
2Y	371	HK-----CRFCAKVGSDSALOIHRSHTGERPKVCNVCGRFTTRGGLKXVHFHRRKXPH	427
2b	205	QKLFTRCQCQGRDESLHLHONTHTGKSTVSCVCGRGFSLKALRLRQTH-----	258
2Y	428	YOMNPHVPEHLVITYSSGLPYGMSVPRPEKAEEBAATPGGVERKPLVASTALSTATES	487
2b	259	-----SG-----EKPLCKVCGRGYT--	274
2Y	488	LHLLSTSAGTATAPGLPAFKFKVLMKAVEPKNKADENTPPGSEBSAISGVASBSTATLMQ	547
2b	275	-----	274
2Y	548	LSKLMTSLPSWALLTNHFS-TGSEPLPL- CARALGASPSFSTSKLOOLEKIDROGAVA	604
2b	275	-----SKSLTYHERHTTGKPYECQCGRPNKSSYNKHLK-----	312
2Y	605	VTSASGAPTTSPAPASSASG--PMOCVITCLRVLSCPRALRLHYGONGGERPFKVC	662
2b	313	-----AHSGEKFVCKECGRGYTNKSYFVHKRIHSGEKPYRCQEC	353
2Y	663	GRAFTSGNLRNRAHFVGHKASPARAQNSCPICQKFTNAVTLOQHYRMHLGGIIPNGTA	722
2b	354	GGFESN-----KSHLTHQRTHSGEKPRACROCKOSFVYKGSILRHQRTHSGEK-----	402
2Y	723	LPEGGAOENGESEOSTV-----SGAGSFPOOOSQOPSPPEELSEEEDEEEDV	775
2b	403	-PFVCKDCERSFQOKSTLYVHQRTHSGEKPFVCRGCGGFIKSTLYVHQITHSEKPFV	461
2Y	776	TDEDSLARGSESGGEKAIVRGDSEASGAEEVGYVAAAATAGKEMDSNEKTQOSSL	835
2b	462	CKD-----CGRGFIKSTFTLHQRTHSEKPYCGRCEG-----RRFRDKSSYKHL	507
2Y	836	PPPPPSDLDQPPMGGSSGVJGK-----DEGKPERSSSPASALTTPGEATSVLYEE	891
2b	508	-----RAHLGEKRFPCRDGR-----GFTLKP	530
2Y	892	LSLOEA-----MRKEPGESSSRKA-----CEVGOAFPSOALEEH	927
2b	531	LTIHQTHSGEKRFMCKQCKSPSLKANILRHQWTHSGERPNCKDCGAGFILLKSTLLFH	590
2Y	928	OKTHKEGRLFTVCFCRQGLERATLKKHMLLAHHVOYF	967
2b	591	OKTHSGEKP-FICSECGQGFIMKSNLVKHQ-LAHSGKOPF	628

RP SEQUENCE FROM N.A.
RC TISSUE=Cerebellum, and Striatum;
RX MEDLINE=97317138; Pubmed=9173996;
RA Loev S.J., Margolis R.L., Young W.S., Li S.-H., Schilling G.,
RA Ashworth R.G., Ross C.A.;
RT "Cloning and expression of the rat atrophin-1 (DRPLA disease gene)
RT homologue.";
RL Neurobiol. Dis. 2:129-138(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain, Cerebellum, Hippocampus, and Substantia nigra;
RX MEDLINE=96081227; Pubmed=8541849;
RA Schmitt I., Epplen J.T., Riess O.;
RT "Predominant neuronal expression of the gene responsible for
RT dentatorubral-pallidolysian atrophy (DRPLA) in rat.";
RL Hum. Mol. Genet. 4:1619-1624(1995).
CC -I- TISSUE SPECIFICITY: PREDOMINANT NEURONAL EXPRESSION, ALTHOUGH
CC MARKEDEX REDUCED AMOUNTS ARE FOUND IN MOST OTHER TISSUES.
CC -I- DEVELOPMENTAL STAGE: SIMILAR EXPRESSION AT ALL DEVELOPMENT STAGES
CC (DAY 14.5 P.C., 17.5 P.C., NEWBORNS AND ADULTS).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; U31777; AA8037.1; -
DR EMBL; X89453; CAA61623.1; -
DR InterPro; IPR002951; Atrophin.
DR Pfam; PF03154; Atrophin-1; 2.
DR PRINTS; PR01222; ATROPHIN.
FT DOMAIN 165 171 POLY-PRO.
FT DOMAIN 303 306 POLY-PRO.
FT DOMAIN 377 383 POLY-SER.
FT DOMAIN 387 391 POLY-SER.
FT DOMAIN 440 446 POLY-PRO.
FT DOMAIN 477 480 POLY-HIS.
FT DOMAIN 481 489 POLY-GLN.
FT DOMAIN 502 505 POLY-PRO.
FT DOMAIN 562 572 POLY-SER.
FT DOMAIN 702 705 POLY-PRO.
FT CONFLICT 455 455 N -> S (IN REF. 2).
FT CONFLICT 594 594 F -> L (IN REF. 2).
FT CONFLICT 689 689 P -> R (IN REF. 2).
FT CONFLICT 717 717 T -> M (IN REF. 2).
FT CONFLICT 737 737 A -> V (IN REF. 2).
FT CONFLICT 965 965 MISSING (IN REF. 2).
SQ SEQUENCE 1183 AA; 124778 MW; 7FB9928DCADF9B1F CRC64;

Query Match 5.0%; Score 265; DB 1; Length 1183;
Best Local Similarity 21.4%; Pred. No. 6.7e-05;
Matches 239; Conservative 94; Mismatches 372; Indels 412; Gaps 52;

OY 42 TDPTETLAHNACTSDPPVAVIIGQENPNNSASSEPRRGGHNNPQVMDTEHSNPDG 101
Db 118 SDPRDI--DQDNSTSPSI-YSPGSVENDSDSSSLSGCAPARYHPPPLPFPSPPPD-- 172
OY 102 SSVPDPTWGPERRGEESGHFLVATGTAA-----GGG 135
Db 173 -SIPROPESGFE-----PHRSVPPTGYHAPMEPTSRLLFGCPPGAPPPHOLYPGSA 224
OY 136 GGIILASPKIG-----ATPLP-----PESTP----- 156
Db 225 GGGVLSGPGMPKRGGAASVSPSGCKQHPPPTPIPISSSGASGAPPAKPPNTPVGAG 284
OY 157 ----APP-----PPP-----GHUNIPILBELRVLQ 189
Db 285 NPSAPPPTAFPHVTPMLPPPALRLPNMASASPPGMGAQPIFGHLPSP----- 333
OY 190 ROIHOMOMTEOICROVLLIGSGTVG-----APASPELPGTGTASTKPLPLFSP 242

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 13, 2003, 15:17:45 : Search time 43.0643 Seconds
(without alignments)
4808.365 Million cell updates/sec

Title: US-09-988-117-1

Perfect score: 5277

Sequence: 1 MAHSESRSLGVAPGEPAAE.....PSITSTGLSPFRDDPTIP 1005

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

- Listing first 100 summaries
- 1: SPREMBL.21:*
 - 2: sp_bacteria:*
 - 3: sp_fungi:*
 - 4: sp_human:*
 - 5: sp_invertebrate:*
 - 6: sp_mammal:*
 - 7: sp_mhc:*
 - 8: sp_organelle:*
 - 9: sp_phage:*
 - 10: sp_plant:*
 - 11: sp_rodent:*
 - 12: sp_virus:*
 - 13: sp_vertebrate:*
 - 14: sp_unclassified:*
 - 15: sp_virus:*
 - 16: sp_bacteriopl:*
 - 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No	Score	Query Match	Length	DB	ID	Description
1	1291	24.5	1330	13	Q91929	Q91929 xenopus lae
2	1280.5	24.3	1261	13	Q902W5	Q902W5 gallus gall
3	1274.5	24.2	1377	13	Q9DDN5	Q9DDN5 xenopus lae
4	1246	23.6	1308	13	Q9DF77	Q9DF77 gallus gall
5	1176.5	22.3	1272	11	Q9EPW7	Q9EPW7 mus musculu
6	1145	20.3	1061	13	Q9PVN7	Q9PVN7 xenopus lae
7	1070	20.3	846	13	Q9PVN6	Q9PVN6 xenopus lae
8	1038	19.7	1053	4	Q9UJQ4	Q9UJQ4 homo sapien
9	1005	19.0	888	13	Q12958	Q12958 oryzae lat
10	910	17.2	382	5	Q9VKH2	Q9VKH2 drosophila
11	784	14.9	1373	13	Q9W7G2	Q9W7G2 gallus gall
12	727	13.8	1263	5	P91639	P91639 drosophila
13	726	13.8	1267	5	Q9VKH3	Q9VKH3 drosophila
14	663.5	12.6	299	13	Q9DGH3	Q9DGH3 brachydanio
15	648.5	12.3	327	13	Q9DGH1	Q9DGH1 brachydanio
16	631	12.0	288	13	Q9DGH2	Q9DGH2 brachydanio

17	513	9.7	549	13	Q9PVN5	Q9PVN5 xenopus lae
18	391.5	7.4	744	5	Q17396	Q17396 caenorhabdi
19	363.5	6.9	2406	4	Q9BXS0	Q9BXS0 homo sapien
20	360.5	6.8	1258	4	Q9P2A7	Q9P2A7 homo sapien
21	346.5	6.6	2282	11	Q61479	Q61479 mus musculu
22	342	6.5	751	4	Q9NS43	Q9NS43 homo sapien
23	334	6.3	927	4	Q9UEG4	Q9UEG4 homo sapien
24	333.5	6.3	619	5	Q96K58	Q96K58 homo sapien
25	333.5	6.3	736	5	Q960L6	Q960L6 drosophila
26	330.5	6.3	1300	4	O15090	O15090 homo sapien
27	329.5	6.2	498	4	Q9H8L4	Q9H8L4 homo sapien
28	324.5	6.1	725	4	Q96S24	Q96S24 homo sapien
29	324.5	6.1	1237	11	Q9D2D7	Q9D2D7 mus musculu
30	322	6.1	1615	13	O57415	O57415 gallus gall
31	321	6.1	734	4	Q96171	Q96171 homo sapien
32	316.5	6.0	726	11	Q92A56	Q92A56 mus musculu
33	316.5	6.0	2232	5	P91365	P91365 caenorhabdi
34	314.5	6.0	962	5	O61360	O61360 drosophila
35	314	6.0	814	11	O70162	O70162 mus musculu
36	313	5.9	1920	5	O46205	O46205 drosophila
37	311.5	5.9	934	5	Q9VKP1	Q9VKP1 drosophila
38	308.5	5.8	1893	5	O9M471	O9M471 drosophila
39	308	5.8	1891	5	O77275	O77275 drosophila
40	307	5.8	1173	13	Q90783	Q90783 gallus gall
41	306	5.8	561	4	Q96BV0	Q96BV0 homo sapien
42	305.5	5.8	1104	4	Q9P243	Q9P243 homo sapien
43	299.5	5.7	700	11	O99K53	O99K53 mus musculu
44	299.5	5.7	701	11	Q9R164	Q9R164 mus musculu
45	299.5	5.7	743	11	O8R5D1	O8R5D1 mus musculu
46	298.5	5.7	744	5	Q9VQ09	Q9VQ09 drosophila
47	298	5.6	553	11	Q921H7	Q921H7 mus musculu
48	298	5.6	591	11	O8R0V0	O8R0V0 mus musculu
49	298	5.6	841	11	O9WVL8	O9WVL8 mus musculu
50	296	5.6	644	4	Q9NT61	Q9NT61 homo sapien
51	295	5.6	496	4	Q96SL8	Q96SL8 homo sapien
52	293.5	5.6	744	5	O24219	O24219 drosophila
53	293	5.6	820	53	Q9Y094	Q9Y094 drosophila
54	293	5.5	884	11	Q99PV8	Q99PV8 mus musculu
55	291	5.5	654	4	Q96CP9	Q96CP9 homo sapien
56	290.5	5.5	691	4	Q9H937	Q9H937 homo sapien
57	290.5	5.5	1636	4	Q9Y474	Q9Y474 homo sapien
58	290	5.5	596	11	Q9CVF3	Q9CVF3 mus musculu
59	288.5	5.5	783	11	Q9QY56	Q9QY56 mus musculu
60	288	5.5	995	11	O35615	O35615 mus musculu
61	287.5	5.4	650	11	O00146	O00146 mus musculu
62	287.5	5.4	784	4	O00146	O00146 mus musculu
63	287	5.4	662	5	Q9N360	Q9N360 caenorhabdi
64	287	5.4	891	5	O61361	O61361 drosophila
65	285.5	5.4	835	4	Q9H165	Q9H165 homo sapien
66	285	5.4	1190	4	Q99621	Q99621 homo sapien
67	284	5.4	599	4	Q9P2E9	Q9P2E9 homo sapien
68	282	5.3	613	4	Q9H785	Q9H785 homo sapien
69	282	5.3	734	11	Q99065	Q99065 mus musculu
70	281.5	5.3	624	11	O62065	O62065 mus musculu
71	281.5	5.3	894	4	Q9COK0	Q9COK0 homo sapien
72	280.5	5.3	786	4	O8TDG8	O8TDG8 homo sapien
73	279.5	5.3	485	4	Q9NRY0	Q9NRY0 homo sapien
74	279.5	5.3	1173	11	Q63624	Q63624 ratu
75	279	5.3	744	11	O62788	O62788 ratu
76	278.5	5.3	618	4	Q96MX3	Q96MX3 homo sapien
77	278	5.3	703	11	Q9R161	Q9R161 mus musculu
78	277.5	5.3	568	11	O8R0T2	O8R0T2 mus musculu
79	277.5	5.3	594	5	P91805	P91805 sarcophaga
80	277	5.2	1054	5	O9VA40	O9VA40 drosophila
81	276.5	5.2	742	4	Q9H0M5	Q9H0M5 homo sapien
82	276	5.2	679	4	O8TF50	O8TF50 homo sapien
83	276	5.2	708	11	Q91XV1	Q91XV1 ratu
84	276	5.2	727	11	O8VC29	O8VC29 mus musculu
85	276	5.2	736	11	Q9RPJ6	Q9RPJ6 ratu
86	275.5	5.2	943	5	O8WRV3	O8WRV3 drosophila
87	275	5.2	556	4	Q9UMP5	Q9UMP5 homo sapien
88	275	5.2	1660	5	Q9XKN0	Q9XKN0 leishmania
89	275	5.2	2703	5	Q9VEG7	Q9VEG7 drosophila

90	275	5.2	2715	5	061603	061603 drosophila
91	274.5	5.2	813	11	09JLIF8	09JLIF8 mus musculu
92	274	5.2	823	4	09H162	09H162 homo sapien
93	273.5	5.2	556	5	09VDN4	09VDN4 drosophila
94	273	5.2	670	4	096E24	096E24 homo sapien
95	273	5.2	671	4	0912A4	0912A4 homo sapien
96	273	5.2	673	4	09UD02	09UD02 homo sapien
97	273	5.2	812	11	09SPV7	09SPV7 mus musculu
98	272.5	5.2	623	11	062510	062510 mus musculu
99	272.5	5.2	688	4	08TA05	08TA05 homo sapien
100	272.5	5.2	756	5	061362	061362 drosophila

ALIGNMENTS

RESULT 1

9191929 PRELIMINARY; PRT; 1350 AA.
C 091929;
T 01-NOV-1996 (TREMBLrel. 01, Created)
T 01-NOV-1996 (TREMBLrel. 01, last sequence update)
T 01-JUN-2002 (TREMBLrel. 21, last annotation update)
DE zinc finger protein.
N XNAL-1.
S Xenopus laevis (African clawed frog).
C Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
C Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
C Xenopodinae; Xenopus.
X NCBI_TaxID=8355;
N [1]
P SEQUENCE FROM N.A.
X MEDLINE=96317243; PubMed=8734496;
X HOLLEMAN T., Schuh R., Piefer T., Stick R.;
T "Xenopus xsal-1, a vertebrate homolog of the region specific homeotic
L gene spalt of Drosophila.";
L Mech. Dev. 55:19-32(1996).
C -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
R EMBL; L46583; AAC4233.1; .
R HSSP; P07248; IARE.
R InterPro; IPR000822; znf_C2H2.
R Pfam; PF00096; zf_C2H2; 10.
R PRINTS; PR00048; ZINC_FINGER.
R ProDom; PD000003; znf_C2H2; 1.
R SMART; SM00355; znf_C2H2; 9.
R PROSITE; PS00028; ZINC_FINGER_C2H2_1; 9.
R PROSITE; PS50157; ZINC_FINGER_C2H2_2; 8.
W DNA-binding: Metal-binding: Nuclear protein; Zinc-finger.
Q SEQUENCE 1350 AA; 145663 MW; B62F60D861C4D2FF CRC64;

Query Match 24.5%; Score 1291; DB 13; Length 1350;
Best local similarity 31.1%; Pred. No. 2, le-67;
Matches 373; Conservativity 163; Mismatches 356; Indels 308; Gaps 44;

Y 8 SSRIGVAGEPRLELGDGA-SEEDHPQVCAKCAQFTPTPELIAHONACSTDPVAVITIG 66
D 38 STRAAAEPRDDDSGNESRSGSEETNVCEKCAAEFFKWTDLHKTKTCKNPLVILVND 97
Y 67 -----QENPNSSASSEPRPEGH-----NNP-----QVMDTE 93
D 98 VAAVVEEVPPESSPASS---PSNHASETAEENIQVENNDICDKDTEKEEPMVEYTE 154
Y 94 HSNPPDSSGSVPTDPTMGPERGERGESSGHELVATGTAAGGGGLIILASPK----- 144
D 155 EKNPSPQASDAPATPL--PQIPESSMTNVMPTNTLTLE-----TLQSTKVAVAOFSON 207
Y 145 ---LGATPLPEESTAPAPPPPPPPPPGVSGHINPLILELKYLOROQHOMQTEBI 201
D 208 AQCVGGVNATAATAT-----MAIPMLIEQLMALDOOQIHQILIDBI 249
Y 202 CROYILLG-----SLGCTVGAPASP-----SELPGGTASTKPL 236
D 250 RSOVALMNRQPLRPPLNPLVPSONAPIPASNOLOGFPAHSHLTQTSVPPILSGPATSG 309

QY 237 IDLFS-----PIKPVOTSKTLASSSSSSSSSGAETPKQAFPHLYH 277
D 310 PPSFENPQHMSOPPSGASPTNIPCPVSSVPTRESTISLSTNSKASSAAPSLSNSTNPH 369
QY 278 PLGSOHPFSAGVGVRSHKPPAPSPALPGSTDLIASPHLAPSPSTGLLAOCLGARGI 337
D 370 PQQSSPTPLSGHNIINSSSSPLPPOS-----SSNSVIFPNPLASIAA---TANAL 420
QY 338 EATASGGLLKPKNGS--GELSYGEVMPLEKPGRGHKKCFKCKVPSDSALQIHRSHTGE 396
D 421 DPLSA--LAKHKKGKPPNVSVFEKTTTSDPEFFKHKKCFKCAKVFSGDSALQIHRSHTGE 478
QY 397 RPKCNVCGNREFTTGNLKVHFRHREKYPHVQNMHPPEHLADVTITSSGGLPYGASVP 456
D 479 RPFKNICGNRSTCKNLKVKHFORHKKRYHIOANPNYPVELDNGPPSSGILPYGASLPP 538
QY 457 EKAEEBAATPGCGVERKPLVASTATLSTLSTASG---TATAPGLPAFNKFVLMK 513
D 539 E-----KPV---TTWILDSKPVLPVPTTIGLQLPPTIGMPGVNSYSDSP 580
QY 514 AVEPKKADEN--TPGSEGSATIS-----GVAESST 542
D 561 SITPSNRSPQRPSPASSECSNLSPEPINNSELCIOASSESPQEPQTRTYTPKQEPYQPS 640
QY 543 ATL-----MOLSKLMT---SLPSWALLTNHRKST-----GSFPLPLCARALGA 582
D 641 SRYVGQRPVNVQSSPYTTPVTPVTDSSVSFHSNSVLPMPMDQKAFKFPFGGLISM-- 698
QY 563 SPSETSKLOOLYEKIDROGAVANTSAASGAPTTAPAPSSASSASPNOCVILRYLSCR 642
D 699 QSSESTKLOOLEVENIDRK-----MDPNOCVIAHRYLSCHS 734
QY 643 ALRLHYGGERPFCKCKVCGRAFTSTGNLRAHFVGHKASPAARONSCPICOKFTNAV 702
D 735 ALKMYRTTGERPFCKCKVCGRAFTTGNLKTGHGVHRSKRPRLRYOHSCPICOKFTNAV 794
QY 703 TLQOHVRLHGOIPNGGTAIPREGGAOENGSEOSTVSGASFPQOOSQOPSPEBELSE 762
D 795 VLQOHRLMHMGQIPN--PRLPEGFONAKDS-----ELSYDOKNLETMSN 837
QY 763 EEEEDDEFEEDVTDSDSLAGRSE-----SGEKATSVRGDSEASGAEEVGYIAAA 817
D 838 YDDDFDNDSEDLDLKDTASDSSKPLIPYSGSSPASSPTVISSIAA--LENMKMIISVM 896
QY 818 TAGK-----EMD--SNEKT-----QOSSLPPEPPPSDLDPQPMPE----- 851
D 897 TAQQTIGLKNENGSEFIDHLSNDSSAVGDLDESQASGAPMSSESSMQVYLSPAHSHSE 956
QY 852 --QGSSGVLGKKEEG---KPERSSPASALPREGA---TSVTLVEELSLOEA--- 897
D 957 SIRSKSPVLISSOEPPVIOLKTEKPDSPIP--TPRENDGVLDLTSNPGRIKEEAPYSL 1014
QY 898 --MKREPGSSSRKACEVCGQAFPSQAALFEHQTHKEGLFLFCVPRQGLERATLTKK 955
D 1015 LFLSRERGFKFS--TYCNIGKPRACKSALFIHRSHKERP--FTCTVCKRGGSYMGNLKO 1072
QY 956 HML---LAHQVQFPAPH--GPONIALSLVPGCSSTITSGLS---PPFRKDDPTIP 1005
D 1073 HLTHTKIKELPQLPEPNTPLGPSQ--TTTSLVTSITAVMIMEVNGHKKPISLGSGPHLP 1131

RESULT 2

Q90ZM5 PRELIMINARY; PRT; 1261 AA.
ID Q90ZM5;
AC Q90ZM5;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, last annotation update)
DE Spalt 2 protein.
GN SAT2.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Archosauroidea, Aves, Neognathae, Galliformes, Phasianidae, Phasianinae;
 OX Gallus.
 RN NCBI_TaxID=9031;
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21184122; PubMed=11287198;
 RA Farrell E.R., Tosh G., Church E., Munsterberg A.E.:
 RT "Cloning and expression of CSAL2, a new member of the spalt gene
 family in chick."
 RL Mech. Dev. 102:227-230(2001).
 DR EMBL: AF304358; AAK38370.1; -.
 DR InterPro: IPR000345; Cyt_c_heme_bind.
 DR InterPro: IPR000822; ZnF_C2H2.
 DR Pfam: PF00096; zf-C2H2; 7.
 DR PROSITE: PS00190; CYTOCHROME_C; UNKNOWN_1.
 DR PROSITE: PS0028; ZINC_FINGER_C2H2_1; UNKNOWN_7.
 DR PROSITE: PS00157; ZINC_FINGER_C2H2_2; 7.
 KW DNA-binding; zinc-finger.
 SO SEQUENCE 1261 AA, 134317 MW, BB2E257BCA8565 CRC64;
 Query Match 24.3%; Score 1280.5; DB 13; Length 1261;
 Best Local Similarity 31.2%; Pred. No. 8.1e-67;
 Matches 366; Conservative 156; Mismatches 362; Indels 289; Gaps 41.

[illegible]

```

RESULT 3
Q9DDN5          PRELIMINARY;      PRT:   1377 AA.
ID Q9DDN5
AC Q9DDN5;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DE 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Spalt transcription factor Sall1.
OS SALL1.
GN Xenopus laevis (African clawed frog).
OC Amphibia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipridae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RA Klingbeil P., Frazzetto G., Boumeester T.;
RT "Sall1, a Xenopus homolog of the human Townes-Brocks syndrome gene
RT SALL1."
RL Submitted (OCT-2000) to the EMBL/Genbank/DDJ databases.
EMBL AK310007; AAC45108.1; -.
DR HSSP; P15822; 1BBO.
DR InterPro; IPR000345; CytC_heme_bind.
DR InterPro; IPR000822; Znf_C2H2.
DR Pfam; PF000096; zf-C2H2; 10.
DR ProDom; PD000003; Znf.C2H2.1.
DR SMART; SM00355; Znf.C2H2; 10.
DR PROSITE; PS00190; CYTOCHROME_C; UNKNOWN_1.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 8.
DR PROSITE; PSS0157; ZINC_FINGER_C2H2_2; 9.
KW DNA-binding; Metal-binding; Zinc-finger.
SQ SEQUENCE 1377 AA; 148861 MW; 68A9FD2B35A16DB CRC64;

Query Match           24.2%; Score 1274.5; DB 13; Length 1377;
Best Local Similarity 31.7%; Pred.No.2e-66;
Matches 359; Conservative 161; Mismatches 336; Indels 277; Gaps
47.

QY 26 SEDHPVCAACAQAQTPTPEFLAHONACSTDPPVVWITIGGOENPNMNSSASSEPRREGHN 85
    :+| |::||::|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db 114 NKEAH--VCNSCCAAEFELSDLLGKHKKNCKTKNLVLIV---NNRPPPAEASSPSFSDN 168

```


15	NPQVM-DT-----EH-----SNPDSGSSVPT	106
169	PDEHMDITINNDOADCDLSKHEHGKPDNEEVEVEISETNNSSGSGNKVYDSSSN	228
107	DPTWG-----PERRGESSGHFLVAATGTACGGGGLILASPRKATPLPPESTPAP	158
229	NPTLGSATTSCLPQVADLTALGNFMSMINSVLIEN-----QSTVVAVAQFSOA---	279
159	PPPPPPPPGVSGHILNIPLILELRVLOQROIHQOMQTEQICROVLLGSLGQTVGAP	218
280	-----KTGGAANNKMLAVAFMEQLALQOOQIHQHLEQIRHQILLAS--QSAEMP	330
219	ASPSELPGCTASTPRLPLFSPI-KPVQTSKILASSSSSSSSSG-----	264
331	TSSSDPQAGLRVSAT-PLTTLSSHLSQLAAAGLAOSIASQASISAGKMLPTIOLPQS	389
265	-----AETPK-QAFHLVPLGSOH-PEFSAVGGRSHKP-----T	297
390	NSVNTVPSICSSSPNNMALATVNPSSERYPSSSTGSOANNQAPAVITSSPPAISLLS	449
298	PAPSPALPGSTDLILASPHLAFPTSTGLLAQCLGAA-----RGLFATASPGI--LKPNG	351
450	PASNPPLPQSPRN-ATP-----ALPSIGTAEADNLSLRKSKPVSYAFETKST	499
352	SGELSYGEVWGLEKQGRKRCRCACAVFGSDSALOHLRSHGTGERPYCNCVGNRPTR	411
500	SDEAFF-----KHKCRCAKAVFGSDSALOHLRSHGTGERPFCNICGNRFSTK	547
412	GNLKHFRHREKYPHVQMPHPVEHLDVITSSGLPYGMSVPERKABEE-----	462
548	GNLKHFRHREKYPHVQMPHPVEHLDVITSSGLPYGMSVPERKABEE-----	462
463	AATPGGVERKPLVASTTALSTESTLLSTAGTATAPG--LRPAFNKVLMAKAVEPKN	519
608	TLTTSYGMLPRLPTLPRTPIKTEEPQIPISHTPASPDSVSKSETASLLKKTIDEPD	667
520	KADEMTPPSEGAISGVASSSTRATIMOLSKLMS-----LPSWALL	561
668	EKEAAMP-----LIDKEHQSQSDSLQNLNTPSACSPTTDSGISVMPNILLP--LM	718
562	TNHFKSTGSPPLPLCARALGASP-SETSKLOQLVEKIDROGAVAVTSAASGAPTSAPAP	620
719	SQGFKA--KEPF--GGILDVTPASERTSKLOQVLENDIK-----	753
621	SSSASGPNQCVICLAVLSCPRALRLHYGHHGERPFCVKYCGRAFTSTGNLRAHFVGR	660
754	---SSDPNECVICHVLSQOSALKMHYRTHTGERPFCVKYCGRAFTTGNLKTHSVHR	809
681	ASPARAONSCPTCKKFTNAVTLQOHVRHMLIGOIPNGTA--LPEGGAQOENG-S	737
810	AMPPLRVQHSCTPCCKKFTNAVTLQOHVRHMLIGOIPNTPVAENYPDSSKS--DTGSFDE	867
738	STVSGAGSPPOOSQOPSPPEELSEEEDEDEEE-----DVTBDSIAGRSSESG	789
868	KTIDLDLN-SDENMDDCPDSSVPDPFKSIDASQDSSSPLPLEVSSITALENOKKLIN	926
790	GEKALSVGRDSEBA-SGAEEVGTVAATAAGKEMDS-----NEKTTQOSSLPP-PP	839
927	AGLADOLASLSAENGVSVDGMTNDSSSLGCDMESQASGPAASESTYSMHALSFPNS	986
840	PPDSLDOPRPMQSGSGVLCGKEGKPERSSSPASALP--EGEATSVTLVEELSLQ	896
987	TIDYIKSPMTDEKLOAV-----SLDPTNGLSPTPANGALDLT--SSNTDK	1031
897	AMRKPE-----GESSRKACEVCGAAPPSSQALTEHOKTHPKEGFLFTCVCFROGF	947
1032	VIKERPLVGLPFPRDRCCKKNTICDLCGKTFACQSALDIHRSHTKER-FICTVCNRRG	1090
948	LERATLKMLLAHNOV-----QPRAPHGPNIALSLVPCGSPSITSTGLSP	995
1091	STKGMLKOHMLT-HQMRDLPLSOLFEP-----SSSMTTP-NPTIPAPSNP	1132

15	NPQVM-DT-----EH-----SNPDSGSSVPT	106
169	PDEHMDITINNDOADCDLSKHEHGKPDNEEVEVEISETNNSSGSGNKVYDSSSN	228
107	DPTWG-----PERRGESSGHFLVAATGTACGGGGLILASPRKATPLPPESTPAP	158
229	NPTLGSATTSCLPQVADLTALGNFMSMINSVLIEN-----QSTVVAVAQFSOA---	279
159	PPPPPPPPGVSGHILNIPLILELRVLOQROIHQOMQTEQICROVLLGSLGQTVGAP	218
280	-----KTGGAANNKMLAVAFMEQLALQOOQIHQHLEQIRHQILLAS--QSAEMP	330
219	ASPSELPGCTASTPRLPLFSPI-KPVQTSKILASSSSSSSSSG-----	264
331	TSSSDPQAGLRVSAT-PLTTLSSHLSQLAAAGLAOSIASQASISAGKMLPTIOLPQS	389
265	-----AETPK-QAFHLVPLGSOH-PEFSAVGGRSHKP-----T	297
390	NSVNTVPSICSSSPNNMALATVNPSSERYPSSSTGSOANNQAPAVITSSPPAISLLS	449
298	PAPSPALPGSTDLILASPHLAFPTSTGLLAQCLGAA-----RGLFATASPGI--LKPNG	351
450	PASNPPLPQSPRN-ATP-----ALPSIGTAEADNLSLRKSKPVSYAFETKST	499
352	SGELSYGEVWGLEKQGRKRCRCACAVFGSDSALOHLRSHGTGERPYCNCVGNRPTR	411
500	SDEAFF-----KHKCRCAKAVFGSDSALOHLRSHGTGERPFCNICGNRFSTK	547
412	GNLKHFRHREKYPHVQMPHPVEHLDVITSSGLPYGMSVPERKABEE-----	462
548	GNLKHFRHREKYPHVQMPHPVEHLDVITSSGLPYGMSVPERKABEE-----	462
463	AATPGGVERKPLVASTTALSTESTLLSTAGTATAPG--LRPAFNKVLMAKAVEPKN	519
608	TLTTSYGMLPRLPTLPRTPIKTEEPQIPISHTPASPDSVSKSETASLLKKTIDEPD	667
520	KADEMTPPSEGAISGVASSSTRATIMOLSKLMS-----LPSWALL	561
668	EKEAAMP-----LIDKEHQSQSDSLQNLNTPSACSPTTDSGISVMPNILLP--LM	718
562	TNHFKSTGSPPLPLCARALGASP-SETSKLOQLVEKIDROGAVAVTSAASGAPTSAPAP	620
719	SQGFKA--KEPF--GGILDVTPASERTSKLOQVLENDIK-----	753
621	SSSASGPNQCVICLAVLSCPRALRLHYGHHGERPFCVKYCGRAFTSTGNLRAHFVGR	660
754	---SSDPNECVICHVLSQOSALKMHYRTHTGERPFCVKYCGRAFTTGNLKTHSVHR	809
681	ASPARAONSCPTCKKFTNAVTLQOHVRHMLIGOIPNGTA--LPEGGAQOENG-S	737
810	AMPPLRVQHSCTPCCKKFTNAVTLQOHVRHMLIGOIPNTPVAENYPDSSKS--DTGSFDE	867
738	STVSGAGSPPOOSQOPSPPEELSEEEDEDEEE-----DVTBDSIAGRSSESG	789
868	KTIDLDLN-SDENMDDCPDSSVPDPFKSIDASQDSSSPLPLEVSSITALENOKKLIN	926
790	GEKALSVGRDSEBA-SGAEEVGTVAATAAGKEMDS-----NEKTTQOSSLPP-PP	839
927	AGLADOLASLSAENGVSVDGMTNDSSSLGCDMESQASGPAASESTYSMHALSFPNS	986
840	PPDSLDOPRPMQSGSGVLCGKEGKPERSSSPASALP--EGEATSVTLVEELSLQ	896
987	TIDYIKSPMTDEKLOAV-----SLDPTNGLSPTPANGALDLT--SSNTDK	1031
897	AMRKPE-----GESSRKACEVCGAAPPSSQALTEHOKTHPKEGFLFTCVCFROGF	947
1032	VIKERPLVGLPFPRDRCCKKNTICDLCGKTFACQSALDIHRSHTKER-FICTVCNRRG	1090
948	LERATLKMLLAHNOV-----QPRAPHGPNIALSLVPCGSPSITSTGLSP	995
1091	STKGMLKOHMLT-HQMRDLPLSOLFEP-----SSSMTTP-NPTIPAPSNP	1132

RESULT 4	
Q9DF77	PRELIMINARY; PRT; 1308 AA.
AC	Q9DF77:
DT	01-MAR-2001 (TREMBLrel. 16, Created)
DT	01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT	01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE	Splatt 1.
GN	SALL.
OS	Gallus gallus (Chicken).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC	Gallus.
OX	NCBI_TaxID=9031;
RN	[1]
RP	SEQUENCE FROM N.A.
RX	MEDLINE=20442162; PubMed=10985862;
RA	Farrell E.R., Munsterberg A.E.;
RT	"Chick splatt 1 gene expression is controlled by a combination of FGF
RT	and wnt signals in developing limb buds."
RL	Dev. Biol. 225:447-458(2000).
DR	EMBL; AF288697; AAG13011.1; .
DR	HSSP; P15822; 1BBO.
DR	InterPro; IPR000345; Cytc_heme_bind.
DR	InterPro; IPR000822; ZnF_C2H2.
DR	Pfam; PF00096; zf-C2H2; 10.
DR	Prodom; PDO00003; ZnF_C2H2; 1.
DR	SMART; SM00355; ZnF_C2H2; 9.
DR	PROSITE; PS00190; CYTOCHROME C; UNKNOWN.1.
DR	PROSITE; PS00028; ZINC_FINGER_C2H2_1; 9.
DR	PROSITE; PS0157; ZINC_FINGER_C2H2_2; 9.
KW	DNA-binding; Zinc-finger.
SQ	SEQUENCE 1308 AA; 139846 MW; 81810CC16336298E CRC64;
Query Match	23.6%; Score 1246; DB 13; Length 1308;
Best Local Similarity	30.7%; Pred. No. 8.9e-65;
Matches 352; Conservative 162; Mismatches 360; Indels 272; Gaps 40;	
Qy	22 GGDASEDHPOVCAKCAQFTDPTFLAHQNAQSTDPVMTVYIGQENPNNSASSEPR- 80
Dy	33 GANRTKNKADHAWGCCAEFFELSDLLQHKKNCTKNQVLV--NENASPETETPPRS 89
Qy	81 PEGHNNPQVMDTEHS-----NPPDSGSSVPTDPTWGPERGESSGHFLVAT 128
Dy	90 PSDNPEQNMNDVTNNMDQVDCSDLSERKRLDKESMDVEASINSSSSSKSVNSITSS 149
Qy	129 GYAAGGGGGLILASPKLG-ATPLPP-----ESTP-APPPEPPPPPPGVSG 173
Dy	150 NSTWKTSAVTTSLPHGDLTTLGNFVYNSNVIIENQSTVVAVAQFQEARCAGNASNN 209
Qy	174 HUNIPLILELRVLOQROIHQOMQTEQICROVLLGSLGQTVGAPASPSELPGTAST 233
Dy	210 KLAVALMEQLALQOOQIHQLEIQIRHQILLAS--QNTDMFTSSSPQGLTRASA- 266
Qy	234 KPLPLFS-----PIRPVQ-----TSKTLASSSSSSSS 261
Dy	267 NPLSTLSSHLSQLAAAGLAQSIASQASISGVKQLPPIQLPQSNPGSTLIPSSSGSSP 326
Qy	262 S-----SGAETPKQAFHLVPLIGSQHPFSAGVGRSHKP-----TPASP 302
Dy	327 NINILAAAYT-----PSSSEKVASISIGSQLSNFPVSASSSPAFALISLLSTASNP 377
Qy	303 ALPGST--DOLIASPHLAFPTSTGLLAQCLGAARGLEATASPGILKPKNGSGELSYGEV 360
Dy	378 ILPQPPPSNISFSP-----LSNIGTPAED-----LNSLTALAQQRKSKPPN---VYAFRA 425
Qy	361 MQLPEGRGHRKRCRCACAVFGSDSALOHLRSHGTGERPYCNCVGNRPTRRGKLVHFR 420
Dy	426 KNSDEAFFKHKRCRCACAVFGSDSALOHLRSHGTGERPFCNICGNRSTKGNLKVHFOR 485
Qy	421 HREKYPHVQMPHPVEHLDVITSSGLPYGMSVPERKABEEAATPGGVERKPLVASTT 480
Dy	486 HREKYPHVQMPHPVEHLDVITSSGLPYGMSVPERKABEEAATPGGVERKPLVASTT 527

OY	481	ALSA	TESTLTLTSTAGTATAPGLPAFNKFLMKAVEP-----	517
Db	528	WLDSK	PVLSTLTSTVGLPLPPTLPSTLPFTKTEBPQPIPIHPSAPSCSVKSDSTADP	587
OY	518	-----	-----KNKADENPPGSGSAISGVAESSTATLMOLSKLM	552
Db	588	TSKIS	NGLSDEVEAGALPTNSCKMEENPQNSSVANNSSSVSPADSSGSSIAFTFN--	645
OY	553	TSLS	WALLTNHFKSTGSPPLPCARLGAASP-SETSKLQOLVEKIDROGAVAVTSAASG	611
Db	646	---PL	MPMSSEQFKA--KEPF---GGLLDSTPASETSKLQOLVENIDKK-----	686
OY	612	APT	TAPAPSSASSSPNCCVLCRLVLSCPRALHVGHGGERPFCKKVCGRAPSTRGN	671
Db	687	-----	-----ADDPNECTICHRVLSQOSALMKMHTHTGERPFCKKTCIGRAFTTKGN	733
OY	672	LRAH	VGHKASPAARAQNSCPICOKKFTNAVTLQOHVHMHLGGQIPNGGTALPEGGAQ	731
Db	734	LKTH	SVHRAMPRLRVQHSCPTCQKKFTNAVTLQOHIRMHMGQIPN--TPVTE-----	785
OY	732	ENG	QSTVYSGAGSFPOQSQSPPEELSEEEDEEEDEEEDVTDEDSLARGSESGE	791
Db	786	--N	PEMESDTSQSFDDKNFDD---LDNFSDENMEDCDPSSVPTPKSADASODSLSSSP	840
OY	792	KAIS	VR-----GDSEASGAEEVGTVA-----AATAGKEMDS-----	825
Db	841	LPLE	MSSTALENOMKMINAGLAEOLOASLKSVENGSVEGDVLTNDSSVGGDMESQASG	900
OY	826	---NE	KTTQOSSLPPPPPSLDOPQPMEOGSSGVLOGKEEGC-KPENSSSPASALTPPE	880
Db	901	SPAV	ESTSMQALSPNSNSTNDYHKSPSTEEKPVRLASSEFANGSLPPRANGSALDLT--	958
OY	881	GEAT	VTLVEELSLQEMAKREPEGSESSRKACEVCGQAFPSQALAEHOKTTHKEGFLTC	940
Db	959	--SS	NDPMKTESLSMLFPPRDRGKFKNTACDTCGTFACOSALDIHYSHKERP-FIC	1016
OY	941	VFC	ROGFLRATATKKNMLLAHNOV-----OPFAPH---GP-ONIALALSVPGCS-PSITS	990
Db	1017	TYCN	GFSTKGNLKOHLMT--KOMRDLPSOLFEPNRSIGPNON---SSVMPANSLSLIK	1071
OY	991	TGL	SPF 996	
Db	1072	TEV	NGF 1077	
RESULT 5				
ID	Q9EPW7	PRELIMINARY;	PRT; 1272 AA.	
AC	Q9EPW7;			
DT	01-MAR-2001	(Tremblrel. 16, Created)		
DT	01-MAR-2001	(Tremblrel. 16, Last sequence update)		
DT	01-JUN-2002	(Tremblrel. 21, Last annotation update)		
DE	Zinc finger protein spalt-3 (Fragment).			
GN	SAL11 OR SAL-3.			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.			
OX	NCBI_TaxId=10090;			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=129/SVEV;			
RX	MEDLINE=21152903; PubMed=11231076;			
RA	Ott T., Parrish M., Bond K., Schwaeger-Nickolenko A., Monaghan A.P.,			
RT	"A new member of the spalt like zinc finger protein family, Msal-3, is			
RT	expressed in the CNS and sites of epithelial/mesenchymal			
RT	interaction.";			
RL	Mech. Dev. 101:203-207(2001).			
DR	EMBL; AF315353; AAG39477.1; -.			
DR	HSSP; P15822; 1BBO.			
DR	MGD; MGI:1889585; Sa111.			
DR	InterPro; IPR000345; CytC_heme_bind.			
DR	InterPro; IPR000822; ZnF_C2H2.			

DR	Pfam:	PF00096; zf-C2H2; 8.	
DR	ProDom:	PDO00003; ZnF_C2H2; 1.	
DR	SMART;	SM00355; ZnF_C2H2; 8.	
DR	PROSITE;	PS00190; CYTOCHROME_C; UNKNOWN_1.	
DR	PROSITE;	PS00028; ZINC_FINGER_C2H2_1; 8.	
DR	PROSITE;	PS00157; ZINC_FINGER_C2H2_2; 8.	
KW	DNA-binding; zinc-finger.		
FT	NON_TER		
SO	SEQUENCE	1272 AA; 134625 MW; 4EB9087166C8BC66 CRC64;	
Query Match 22.3%; Score 1176.5; DB 11; Length 1272;			
Best Local Similarity 30.9%; Pred. No. 1e-60;			
Matches 363; Conservative 142; Mismatches 339; Indels 329; Gaps 50;			
OY	40	QPTDPEFLAHQACSTDPVYVYIIGQENNNSSASSEPREGHNNPQVMDREHNPDP	99
Db	1	EEFEVSDLLHKRSCTKNQVLV--NESPASPKATPPGCP-SLMDPDQMKDAANKAN	56
OY	100	SGSSVPTDPTWGPERRGESESSGHFLVAATGTAAGGGG-----	137
Db	57	QEDCSDLSEPKGLDRGSMEEVEVPATTTTGTGGSGSGSTLSGVNTITPPSCHSGSS	116
OY	138	-----LILASPKLG-ATPLPP-----ESTP-APPPPPPPPPPGVSGHLNT	177
Db	117	TGTSAITTSIPQLGDLTTLTGNFSVINSVNIENLQSTKVAVAQFSQAKRCGASGGKILI	176
OY	178	PLIEELRVLQOQHOMQTEQICROVLLGSLGQTVGAPASPSLPGTGT-ASSTKPL	236
Db	177	STLMEQLALQOQDHLQDLQKHLQHLILAS--QSANLPAPAP-IPSGGLRTISANPL	233
OY	237	LPLFSPIK-----PVQTSKTLASS-----SSSSSS-----SGAETPKQA	271
Db	234	TTLSHLSQQLAAVAGLAQSLASQSANISGVKQLPHVQLPQSSSGTSIYPPSGTSPNMS	293
OY	272	FPHLYPLGSHPPSAGGVGRSHKPTPA-----PSPA-----LPGSTDQILASP--HLA	318
Db	294	IYTAAVPTPSESEKY-ASNAGASHVSPVAVSASSPAFALISLSLPSNPRLPPTPANAV	352
OY	319	FPSTTGILTA--AOCIGARGLEATASPELLKPKNSGELSYGEVMQRLKPGGRHKRCFC	376
Db	353	FPTPLPLNIAATTAEDLNS---LSALAOQKRSKPPN-----VTAEPAKSTDEAFKKHKRCFC	405
OY	377	AKVFGSDSALQIHLRSHTGERPYKNCNCGNRTTNGNLKVHHRHREKYPHYVOMNPPVP	436
Db	406	AKSFGSDSALQIHLRSHTGERPYKNCNCGNRTTNGNLKVHHRHREKYPHYVOMNPPVP	465
OY	437	EHLDYITSSGLPYGMSVPPPKAEAEATPGGCVERRKPLVASTTALSATESLTLSTSAG	496
Db	466	EHLDDNVPTSGIPYGMISIPPE-----KPV---TSWLDYTKRVLPLTLTTSYVG	507
OY	497	TATAPGLPAFNKFLMKAVEP-----KNKADENMP-----PGS-----EG	531
Db	508	LPLPTPLSLTPFI--KTEERAPRISHSASAPQGSVKSQSDGAPDLATRNPSGVPEEVBG	565
OY	532	SAI-----SGVASSSTATLMQSLMTSL-----PSWALLTNHFKSTGS	570
Db	566	SAYPPFGKGEBSMMASAVPTAGN-STLNSPVADGPGCTPTFKPLPLMSSEQFKA--K	622
OY	571	FPLPLCARALG-----ASPSETSKLQOLVEKIDROGAVAVTSAASGAPTTSAAPSSAS	625
Db	623	FP-----FGGLDFQAQASSETSKLQOLVENIDKK-----A	651
OY	626	SGNOCVTCRLVLSCPRLRLHYGHHGGERPKCKVCGRAPSTRGNLRAHFVGHKASPA	685
Db	652	TDPNECTICHRVLSQOSALMKMHTHTGERPRPKKTCISADFTTKGNLKTIVSVHRAMPPL	711
OY	686	RAONSCPICOKKFTNAVTLQOHVHMHLGGQIPNGGTALPEGGAQENGSEOSTVSAGS	745
Db	712	RVQHSQCPICQKKFTNA-CLPAAYRLHMGSQLPN--TPVPD-----NYPEMESDTS	760
OY	746	FPOQSQSPPEELSEEEDEEEDVYTEDDSLARGSESGEKAISYVRGDSSEASG	805
Db	761	F-----DEKNFDDLDNFSDENME-----CPGSGISPTDKSADASQDSLSSSP	803

[illegible]

QY	807	EEEVGTVAAAAATGAKEMDSNEKTTQSSLLPPPPPPDSDLDQPPMNGSSGSSVLGKRECGK	866
		: : : : : : : : : : : :	
Db	740	EXPTKSPROSEARTEPSPINSQSDSNAS-KSPSPYNGD-----DLGML-----SK	784
QY	867	PERSSSPASALPEGE-ATSVTLVEELSLQAMKREP	903
		: : : : : : : : : : : :	
Db	785	DEHSON--GSLNPDGQALDILT---NGFAKKIKKEP	817
RESULT 8			
	Q9U04	PRELIMINARY; PRT: 1053 AA.	
AC	Q9U04		
DR	01-MAY-2000	(TREMBLrel. 13, Created)	
DT	01-MAY-2000	(TREMBLrel. 13, Last sequence update)	
DE	01-JUN-2002	(TREMBLrel. 21, Last annotation update)	
GN	DJ1121P9.1	(novel protein similar to SALL1 (sal (Drosophila)-like 1) (LOC51767)).	
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RA	Sequence from N.A.		
RA	Babbage A.;		
DR	Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.		
DR	EMBL: AL034420; CAB61485.1;		
DR	InterPro: IPR008822; Znf.C2H2.		
DR	Pfam: PF000096; Zf.C2H2; 8.		
DR	ProDom: PD000003; Znf.C2H2; 1.		
DR	SMART: SM00355; Znf.C2H2; 7.		
DR	PROSITE: PS00028; ZINC_FINGER_C2H2_1; 7.		
DR	PROSITE: PS50157; ZINC_FINGER_C2H2_2; 7.		
DR	DNA-binding; Metal-binding; Zinc-finger.		
SQ	SEQUENCE 1053 AA; 112231 MW; 61DD01F21CB2B337 CRC64;		
Query Match			
	Best local similarity	19.7%; Score 1038; DB 4; Length 1053;	
	Matches 317; Conservative 124; Mismatches 328; Indels 320; Gaps 38		
QY	8	SSRLGVAGGEAELGGD--ASPE-----DHPOVCAKCAQFTPTPEFLAHONACST	56
		: : : : : : : : : : : :	
Db	40	AGELGAPVNH--GNDEVASEDEATVKRLREETHVCEKCAFEFISSEPLEHKKNCTK	96
QY	57	DPPLVVIIGQENP-----NNSASSEPRPEG-----HNNPOVMDTEHSNPPDGSV--	104
		: : : : : : : : : : : :	
Db	97	NPLVL-ITNDSGPRPSEDFSGAVLSHQPSPSKDCHRENGSSSEDMKEREPAESVYL	155
QY	105	-----PTDPTWGPERGSESSGHFLVAATGTAAAGGGGLILASPKIGALPPLPESTPA	157
		: : : : : : : : : : : :	
Db	156	KTEPLPTPODISYLAGKVAANTVNTLQA-----LKTGYAVNQRSADALPA	203
QY	158	PPPPPPPPPPGVSGHINIPILTEELRYLQORQIHOMOMTEQICRQVLLGSLGQTWGA	217
		: : : : : : : : : : : :	
Db	204	PVP-----GANSIPVLEQILCLQQOOLQOILQLEQIRIYVNMAS	244
QY	218	PASPELPGTGAASSTKPLPLPSPKIPQVTSKTIASSSSSSSSSSGAETPKQAEFFHLXH	277
		: : : : : : : : : : : :	
Db	245	-----HALHSSGGA-----DTLKLIGSHMSQOVSAVA-----	273
QY	278	PLSGHPHPSAGVSGRSHKPTAPSPALPGSTDOLIASPHLA-----FPSTTGLL	326
		: : : : : : : : : : : :	
Db	274	-LISQKASOGSLDALKQAKLPHANIPSPATSL--SGGLAPFTLKGDTGTVLPNVM SRL	330
QY	327	AAQCGAARGLEATASR-----GLLKPRNGSGE--LSYGEVMPGLEKPGRRKRCFCANV	379
		: : : : : : : : : : : :	
Db	331	PSALLPQAPGSLVLPQSEFSTVALDTSKKGKGPKNISAVDYKPKRBAALYKKKKCKCSNV	390
QY	380	FGSDALQILHLSHTGERPKVCNVCGNRFTTRGNLKVHFNHREKYPIVOMNPHVPEHL	439
		: : : : : : : : : : : :	
Db	391	FGTDSLQILHLSHTGERPVCNVCGNRFTTKGNLKVHFNH-----POVKANDPLFAEQ	446


```

2Y 440 DYLVTSSGLPQMSVPEPEKAEAEAPFGGVERKPIVASTALSTLESITLSTSGAT 499
2b 447 DKVANGNGTLPALVDPDPE-----PSLSIDSKPVLVTTV----- 483
2Y 500 APGLPAPFNKFLVLMKAAVEPKNAKADENTP-----PQSEGS-AISGVASSPATLMQLSK 550
2b 484 -GLPQ-----NLSSGNTNPDLRGSLPQDLPQPSPEBEGPTLPVGCPNTYSP----- 531
2Y 551 LMTSLPSMALTLNHRKSTGSPFLPLCARALGASPESTSKIQLVKIDROGAAVATSAAS 610
2b 532 -----RAGGPDGSGT-PEP-----GSETLKLQOLVENIDRK----- 560
2Y 611 GAPTPSAPAPSSASGSPNOCVLCRLVSCPRALRLHYGHGSGEBREKCYGGRASFTRG 670
2b 561 -----ATIDPECLICHRYLSCQSSSLKMHYRTHTGERPFQCKICGRAFTSG 607
2Y 671 NLRAHFVGHKASPAARAQNSCPICQKKTFTNAVTLQGHVRMLGQIPNGCTALPEGGAA 730
2b 608 NLKTHLGVHRTWTSTIKTHSCPIQCKKFTNAVMLOQHRIHMGQOIRP--ITLPE--NPC 663
2Y 731 QENSGEOSTVSAGS-----FPQOQSOQPSPEEL----- 760
2b 664 DFTSGSEPTVGBNGSTGALCHDDVIESIDVEEVSQGEAPSSSKVPTPLPSIHASPTLG 723
2Y 761 -----SEEEEDDEEEDVDEDESLARGSSEGEKATSVGD 799
2b 724 FAMMASIDAPKCVGAPFPLQROGSGENGSVEEDGLNDS-----SLMGD 769
2Y 800 SEEASGAEEVGTVAATAATGKEMDSNEKTTQGSLSLPPPPSLDOPQPMDOGSSGYLG 859
2b 770 QEYQS-----RSPDILETTTSFQALSPANQSAESIKSKSP---DAGSA 809
2Y 860 GKKEGGKPE---RSSPSASALPRREGATSV---TLVEELSLQAM---KKEGESSSR 908
2b 810 ESSENSKREMEGRSSLPSTFTIAPPTVYAVEVPGTFVGPSTLSPGMPTLLAOPRQAKQ 869
2Y 909 KACEVCGQAFPSQALAEHQKTHPKKGLFTFCVFCROGFLERATIKKMLLAHQVQFPA 968
2b 870 HGCTRRCGKNFSSASALQHERHTGGERP-FVCNICGAFITTKGNLKYH-----YM 918
2Y 969 PHGPONTAA 977
2b 919 THGANNNSA 927

:RESULT 9
:12958 PRELIMINARY; PRT; 898 AA.
D 012958
IC 012958;
2Y 01-JUL-1997 (TREMBLrel. 04, Created)
2b 01-JUL-1997 (TREMBLrel. 04, last sequence update)
2Y 01-MAR-2002 (TREMBLrel. 20, last annotation update)
2b
2Y zinc finger protein sal (Fragment).
2b
2Y Oryzias latipes (Medaka fish).
2Y Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
2Y Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
2Y Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;
2Y Belontiiformes; Adrianichthyidae; Oryziinae; Oryzias.
2Y NCBI_Taxid=8090;
2b
2Y [1]
2Y SEQUENCE FROM N.A.
2Y Stick R., Koester R., Witbrodt J.;
2Y "Hedgehog signaling activates spalt at the midbrain-hindbrain boundary
2Y of fish."
2Y
2Y Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
2Y EMBL: U77376; AAB51127.1; -.
2Y HSSP: P15822; 1BBO.
2Y InterPro: IPR000822; ZnF_C2H2.
2Y Pfam: PF00096; zf_C2H2_7.
2Y SMART: SM00355; ZnF_C2H2_7.
2Y PROSITE: PS00028; ZINC_FINGER_C2H2_1; 7.
2Y PROSITE: PS0157; ZINC_FINGER_C2H2_2; 7.
2Y

```

QY	337	LEATAS-----PGLTKRNGS-GELSTGEVNGRPLEKPGGRHKRCFAKVFQSDALQIHL	390
QY	337	LEATAS-----PGLTKRNGS-GELSTGEVNGRPLEKPGGRHKRCFAKVFQSDALQIHL	390
Db	6	IATVANAALDPLSALMLKHKRKPKNVSVDFDKPSSEDPFKHKRCFAKVFQSDALQIHL	65
QY	391	RSHTEGRRYKCNVGNCFETTKGKLVHFRHREKXYDVMONPPRPHLYVTYSSGLPY	450
Db	66	RSHTEGRRYKCNVGNCFETTKGKLVHFRHREKXYDVMONPPRPHLYVTYSSGLPY	125
QY	451	GMSVPPPEKAEAEATPGGVERKPLVASTALASSTLSTASGTATAPGLPA-FNKF	509
Db	126	GMSVPPPEKAEAEATPGGVERKPLVASTALASSTLSTASGTATAPGLPA-FNKF	167
QY	510	VLMKAVEPKMADE-NTPPGSEGAIS-GVAMSSAT	551
Db	168	AESPSTLPLRSRPNRHPSPSESCASLSNVNALDSTRITTSPPSPNVLGSDGPPILKEGI	227
QY	552	MISLPSMA-----LITNHFSTG-----SFP-L	573
Db	228	LIS-PNYSARPGENTTTTTVTVOLSTNTSTSSGQVSESISSPSASNAVHPVL	286
QY	574	PLCARALGA-----SPSESKIQOLVERIDROGAVAVTSAASGAPTTSAAPSS	622
Db	287	PLMSEOPKAFPRPGGLDMSMOTSETSKIQOLVENTDK-----	324
QY	623	SASSGNOCVILCLRVLSCPRALRLHYGNGEREPKCKYCGRAFTSGNLRHAFVGHKAS	682
Db	325	--MTDNOCVICHRLVSCOSALKMHYRLHTEGPREPKCKICGRAFTTKGNLKTGFVHRSK	382
QY	683	PARAONSCPIQCKKFTFNATVLOOHVNRHLGQITPNCGTALREBGGAAQENGSEOSTVSG	742
Db	383	PLRVHOSCTCCKKFTFNNAVVLQOHIRNHMGQIPN-TPLE-SIQEMETDL-----	432
QY	743	AGSFPOOQSQRPPEELSEEEDEEEDVDTE-----	778
Db	433	--SEDEKSLDAMSNYDDDLIDEMQANDDEEDLKEGELDPKRPYSGSSPTMISSIAA	490
QY	779	-----DSLAC-----RSGSEGE-----KAISRGDSF-----EA	803
Db	491	MENQMKIDSTAMNTHSFGCKPAONGSSFGEGADCFETDLSLAVGAEQOSLSPALSES	550
QY	804	SGAEEEV-----GTVAATAATACKEMDSNEKTKTOOSSLP PPPP	842
Db	551	SGSMOHLSTAHSHSEOSRKSPPALNNNNNSISMVYEEQOENNTAGITITYKSESETPSP	610
QY	843	S-----LDOPDM-----EQSSGVILGKEEG-GKPERSSSPASALPDE-GEAT	884
Db	611	SAIEGTGALDLTATQPSRHYIKESHFSMFLNDRGLSAPVILASTASNMIKEMNGHCK	670
QY	885	SAVTLVEELSLQEAAMR-----KEGGESSSRKACEVCGCAFPPQAALEENQ	928
Db	671	SLSDNHHLGIQVPAAPPTTMSPIINDMLAPPPRPTPKQHNQSCGKNFSSASALOIEH	730
QY	929	KTHPKRGPLTCVCFQSGELERATLTKKHM	957
Db	731	RTHTEKRP-FACISICGRAFTTKGNLKVHM	758


```
3N SAL.
3S Gallus gallus (Chicken).
3C Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
3C Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
3C Gallus.
3C NCBI_TaxID=9031.
3X
3X SEQUENCE FROM N.A.
3X Reveal J.-P., Thaller C., Eichele G.;
3X "Evidence for morphogenetic signaling by BMP-2 in the chick limb
3X bud."
3X Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
3X EMBL: AF110143; AAD3404.1;
3X InterPro: IPR000822; znf_C2H2.
3X SMART: SM00355; znf_C2H2; 5.
3X PROSITE: PS00028; ZINC_FINGER_C2H2_1; 5.
3X PROSITE: PS0157; ZINC_FINGER_C2H2_2; 5.
3X DNA-binding; Metal-binding; Zinc-finger.
3X
3X NON_TER 1
3X FT 382 382
3X SEQUENCE 382 AA; 41778 MW; B66226E3DEFDE3B6 CRC64;

Query Match 17.2%; Score 910; DB 13; Length 382;
Best Local Similarity 47.7%; Pred. No. 9.9e-46;
Matches 194; Conservative 53; Mismatches 92; Indels 68; Gaps 12;

347 KPKNGSGELSYGVWGPLEKPGGRHKRCFCAYVGSDSALQIHLRSHTEGRRPKYKNCN 406
3X |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3X 8 KPPN-----VSVEPRSSSDPRFKHKRCFCAYVGSDSALQIHLRSHTEGRRPKYKNCN 63
3X
3X 407 RFTTRGNLKVHFRHREKRYHVNOMRPHVREHLDYVITSSGLPGMSVPPER-----AEE 461
3X |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3X 64 RFTSGKNLKVHFRHREKRYHVNOMRPHVREHLDYVITSSGLPGMSVPPER-----AEE 461
3X
3X 462 EAAPR-----GGVERKPLVASTALSTESTLTLSTS-----AG--TATPGLPAFVK 508
3X |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3X 124 KPVLPPTPTSTIGLDPPTLPVGVSYGDSPTTPRSRSPQRSPASGECTSLSPSLNTSES 183
3X
3X 509 FVLKMAVEPR-----NKADE--NTPPGSE-----GSAISGVAESSTATLMQLSKLM 552
3X |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3X 184 GVPVSAESPPOVSSSVTKAEPIELNPASTRLGDHSLGGQVSTASTSIPITVTDSSVS 243
3X
3X 553 TSLPSMAL--LRNHEFSTGFFRLPCARALGASPESETKLQDLVEKIDGAGAVVTSAS 610
3X |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3X 244 TSLPVPVLRVAVSDQFKA--KPPGGLLDSM--QTSEYSKLQDLVEIDTKK----- 289
3X
3X 611 GAPTTSAFAPSSASSGPNQVCICLRVLSCPRALRLHYGOHGERPFKCVGSRFSTRG 670
3X |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3X 290 -----MTDPNQCVCICLRVLSGOSALKMHTRTHTGGERPFKCKICGRAFTTKG 335
3X
3X 671 NLRANFVGHKASPAARQNSCPTCKKFTNAVTLQAHVAMILGQIP 717
3X |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3X 336 NKTHTGVHRAKRPRLRVQHSCTCKKFTNAVTLQAHVAMILGQIP 382
3X

RESULT 11
ID Q9VVKH2 PRELIMINARY; PRT; 1373 AA.
Q9VVKH2 AC
Q9VVKH2 01-MAY-2000 (TREMBLrel. 13, Created)
Q9VVKH2 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
Q9VVKH2 01-MAY-2002 (TREMBLrel. 20, Last annotation update)
Q9VVKH2 SALM OR CG6464.
Q9VVKH2 GN
Q9VVKH2 Drosophila melanogaster (Fruit fly).
Q9VVKH2 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Q9VVKH2 Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Q9VVKH2 Ephydroidea; Drosophilidae; Drosophila.
Q9VVKH2 NCBI_TaxID=7227;
Q9VVKH2 [1]
Q9VVKH2 SEQUENCE FROM N.A.
Q9VVKH2 RP
Q9VVKH2 STRAIN=BERKELEY;
```

```
3X MEDLINE=20196006; PubMed=10731132;
3X Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
3X Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
3X George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
3X Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
3X Brandon R.C., Rogers Y.-H.C., Blaise J.R.G., Champe M., Pfeiffer B.D.,
3X Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
3X Abriti J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Basley E.M.,
3X Bailew R.M., Basu A., Baxendale J., Bayraktaroglu L., Bolashkov S.,
3X Beeson K.Y., Benos P.V., Bertram B.P., Bhandari D., Bolshakov S.,
3X Borokova D., Botchan M.R., Bouck J., Brockschtein P., Brotler P.,
3X Burlis K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
3X Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
3X de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
3X Doonan K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
3X Durbin K.J., Evangelista C.C., Ferraz C., Ferrier S., Fleischmann W.,
3X Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
3X Glodex A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
3X Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
3X Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwan C.,
3X Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
3X Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
3X Laoko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
3X Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
3X Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
3X Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
3X Nelson D.R., Nelson K.A., Nixon K., Pollard J., Puri V., Reese M.G.,
3X Palazolo M., Peltman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
3X Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
3X Shue B.C., Siden-Kiamos I., Simpson M., Skupski P., Smith T.,
3X Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
3X Sylvestre R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
3X Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
3X Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
3X Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
3X Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
3X Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
3X "The genome sequence of Drosophila melanogaster."
3X Science 287:2185-2195(2000).
3X
3X EMBL: AE003632; AAP53097.1;
3X HSSP: P15822; 18B0.
3X Flybase: FBgn0004579; salm.
3X InterPro: IPR000822; znf_C2H2.
3X Pfam: PF00096; znf_C2H2; 7.
3X SMART: SM00355; znf_C2H2; 7.
3X PROSITE: PS00028; ZINC_FINGER_C2H2_1; 7.
3X PROSITE: PS0157; ZINC_FINGER_C2H2_2; 7.
3X DNA-binding; Metal-binding; Zinc-finger.
3X
3X SEQUENCE 1373 AA; 151014 MW; A6D857870F645ACC CRC64;

Query Match 14.9%; Score 784; DB 5; Length 1373;
Best Local Similarity 23.6%; Pred. No. 1.1e-37;
Matches 310; Conservative 144; Mismatches 421; Indels 440; Gaps 46;

68 ENPNNSASSEPREGHNNPVMDTSHN--PPDSGSSVPTDPT-----WGPE 113
3X |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3X 106 ENDIKSEAKSEIEVEDNNNNVAMTKPSEEREPRNAGSMSPSSVVAESAEEATERPTE 165
3X
3X 114 RREGE-----ESSGHFLVATGTAAGGGGLI-----LASPKLGATPLPE 153
3X |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3X 166 KEREKDEVDVEKDEDAFSSAVPSTEVTLPGAGAVTLEAIOMQMAIQFAKTIANG 225
3X
3X 154 STFAPPPPPPPPPPGVSGHLNPLILEELRVLAQRIHOMQTEQICROVLL----- 207
3X |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3X 226 SNGADNE-----AAKKQLAFLOOTLNFNLQOQOLFQIOLQLOLQALNOAKQE 274
3X
3X 208 -----LGSIGQVGAAPSPBELGTGTAASSKRLPLPLFSP 242
3X |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3X 275 EDTEEDADQEDODQETDTYEERETADMLRQAKARMAE-----AKAKOHILMAGVP 328
3X
3X 243 IKPVQTKTLASSSSSSSSSGAETPKQAFHLVHPLGSGHPFAGVGRSHKPTPA--- 299
3X |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3X 329 LR-----ESSGPAESLKRREHHDHESOPNNRP-----SLDNTHKADYADDA 370
```



```

300 -----PSPALPGSTD--OLIASHPH--LAPSTTGLLAQCSGAARGLFATASPGILKP 348
371 LAKKEMETPRLPFGSDLASITNNHDLRPNNSLDLQKR-----AOEVLDSAGQIL-A 425
349 KNGSGELSTGYEWGP---LEKPGGRHKCRFCAKVGSDALSALQILHNSHTGERPKCVNG 405
426 NSMADDFAVGEKSGEKKGRMEPEFKHRCRCYSGKVGSDALSALQILHNSHTGERPKCVNG 485
406 NRETFGNLKVHNRHREKRYRHYOMNHRVREN-----LDYVTTSSCLPRGMSVRE 457
486 SRETTGNLKVHNRHREKRYRHYOMNHRVREN-----LDYVTTSSCLPRGMSVRE 545
458 KAEEDAAAPRG--GVER--KPLVASTALASSTESLTLTSAQT--ATAPGLPAFKFV 510
546 LGSAPASFPAPRFGILNLYKRRPMEILKSIGAAARHQYFQDELPTDLKRPSPQDLEDERQV 605
511 LMKAVERNKADENTPRGSGSAIS-----GVAESSTATLMOLSK-----LMTSL 555
606 KNEPVEEKDQREHEQEMAECSSEPERPLPLEVRKKEKREVEQEQVEDHRIEPRRTPS 665
556 PSMALLT--NHFKSTGSP--LPLCARAL--GASPSSTSLQOL----- 593
666 PSSEHRSFNHHHSHNGYRVVQVPIQRAALMHPOSSQSHLDHLPTPGQLPRDEFEA 725
594 -----VEKIDROGAVALVTSASGAPTTSAAP 620
726 ERFPLNFTTAKMLSEPHNSRVNSRAGALPRGVPRPNNHNMARSPEFNRIKHEAAL 785
621 -----SSSASGPNOCVTCILRVLSCPRALRYLHYQ 650
786 LPRHNSNDSMNFLEVSTNTECETMKLAKELMKNKKTISDNOCVSCRVLSCKSKSLQMIHYRT 845
651 HGERPERKCVKCGRAFTSTGNLKAHNVGNKASPARAKONSCPTCOKKFTNAYTLQOHVRA 710
846 HTGERPRFCRIGRAFTTKGNLKTMAVHKIRPRMRFHQCRKKTYSNALVLOQHRL 905
711 HLG-----GQIPNGTALPRGGGAQDE 732
906 HTGERPTDLPQIOAETRDPRPSMMPGHEMNPRAAAGFNRGALP--GGGPRGPRMNGAH 964
733 NGSEOSTVSGAGSFPOQSOQSPREEELSEE-----EEFEDEEEED----- 774
965 NGALGSE--SSGGMDMDNDGSDYDDVDSSEHLSNMLEQGDGDSRSGDEKSLFEQKL 1023
775 -----VTDESLAGSGESGEKA---ISVRGDE---EASGAEVEVG--- 811
1024 RIDATGVVNTNPRVPRSSASSHSHVSTSAPTSFSVHASQVILKRSSPARSASQAL 1083
812 --TVAATATACKEMDS--NEKTTOOSSLP-----PPRPDLSL----- 845
1084 DLTFRRAAPRTSSSSSRSPRKERVBPSPSLRPSGSSSHASANILTSLPLRPVIGIDLPRG 1143
846 -----QOPRMEGSSGVLG----- 860
1144 LQNHLLQOHNLMOQAAVAAAHAAQHNNHQAALQHOEQLREAAEOKAAAAA 1203
861 -----KEBGG-----KEPRSSPASALPDE---GEATSVTLVEE 891
1204 AAAAAAAROTPRARDROBEGRGAGRPNNPLMGARPRGFMFNPLPLFPRATTONMCA 1263
892 LS--LOEAMKRP-----GESSSRKACEVCGQARFPOQALEHOKTHREGLFTCYF 942
1264 MNOIASVMAAPRNPRLALSGVRGS--TTCGICVKTFFCHSALEIHNHSHTERP--FKCSI 1321
943 CROGFLERATILKHKMLLANHVOQPARPHQNRQIAIALSLVPGSPSITSTGSPFR 997
1322 CDRGFETTKMLKQHMILT--HKIRDMEOETFRNRAY-----KYVSLPISPIR 1365

```

SUIT 12
1639
P91639 PRELIMINARY; PRT; 1263 AA.

```

AC p91639;
DT 01-MAY-1997 (TREMBLrel, 03, Created)
DT 01-MAY-1997 (TREMBLrel, 03, last sequence update)
DT 01-MAY-2002 (TREMBLrel, 20, last annotation update)
DE Zinc-finger protein SALR (SPALT-related protein).
GN SALR OR CG4881.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RA SWRAIN-OREGON-R;
RA Barrio R., Shea M.J., Garulli J., Lipkow K., Gaul U., Frommer G.,
RA Schuh R., Jackle H., Kafatos F.C.;
RT "The spalt-related gene of Drosophila melanogaster is a member of an
RT ancient gene family, defined by the adjacent, region-specific homeotic
RT gene spalt."
RL Dev. Genes Evol. 206:315-325(1996).
CC 1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
EMBL: Y07653; CA68937.1; -.
DR HSSP; P15822; 1BBO.
DR FlyBase; FBgn0000287; salr.
DR InterPro; IPR003006; IG_MHC.
DR InterPro; IPR000822; ZnF_C2H2.
DR Pfam; PF00096; zf-C2H2; 8.
DR PRINTS; PR00048; ZINC_FINGER.
DR SMART; SM00355; ZnF_C2H2; 8.
DR PROSITE; PS00290; IG_MHC; UNKNOWN_1.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 8.
DR PROSITE; PS0157; ZINC_FINGER_C2H2_2; 7.
SQ DNA-binding; Metal-binding; Nuclear protein; Zinc-finger.
KW SEQUENCE 1263 AA; 138991 MW; 2025B3BE7417622 CRC64;

Query Match 13.8%; Score 727; DB 5; Length 1263;
Best local similarity 23.5%; Pred. No. 2,1e-34;
Matches 275; Conservative 144; Mismatches 397; Indels 356; Gaps 39;

QY 67 QENPNNSASSSEPRREGNNNQVMOTENSNRPDSGSSVPDPDTGPERKGESSGHF--- 123
DB 38 RENDRDTLLKELDADSDNSNGTEPQMEAAVBPESDTERETAEEEGEEDPEPNSNALDLS 97
QY 124 LVATGTAAGGGGGLI-----LASPKLGATPLPRESTPAPRPPPPPPPPPPVSGSH 174
DB 98 LISSGGRSLPGSGHVSLEALQHTKVAVAQFATMAAGNQA----- 140
QY 175 LNIPLTLEELRVLOQROITHOMOTEOICROV-----LLLG--SLGQTVGAPASSELPGT 227
DB 141 -DIAMVOSTIFNVOROHIMQLOLQILOLSQIKRAEAAALGRHSHSDEEERPEPEPK 199
QY 228 GTASSTKPLPLFSPKRVQVTSKTLASSSSSSSSSSSGAETPKQAFPHLHNLGSHPS- 286
DB 200 QPTNGLKEELELEGGPSEQDHEH--RREENSKTDKRGTEBRKA-----EEGYQSMKCD 252
QY 287 -----AGVGRSHKTPAPS-----PALPGSTDOLIASPHLAPSTTGLLAQCSGAAR 336
DB 253 ISSSLASSITNNHDPAPRPNPCLEMLQRRNEVLDS-----ASQILHAAQ- 299
QY 337 LEATAPRGLKPKNGSELSTGYEWGPRLEKPGGRHKCRFCAKVGSDALSALQILHNSHTGE 396
DB 300 MOEEYSEYASKEAOSREIF-----KHCKKCVGKIFGYSALQILHNSHTGE 346
QY 397 RPYKCVNCGNREFTTGNLKVHNRHREKRYRHYOMNHRVRENLDYVTTSSG----- 447
DB 347 RFYCVNCGSKFTTKGNLKVHNYQHTQILFPPMLRPGVAPN-----VGHSGQGVQVGEQY 401
QY 448 ---LRYGMSVPRKEA-----EEAATPGGVERKPLVASTALASATESLTLTSTA 495
DB 402 PIRLPFAPVAPVAGQEOHONQVEEELIPVQAEDSLKRVAKKE--KSHSPVCEVKTPK 459
QY 496 GTATAPAGLPAFNKFLVLMKANVERPKKADENTPRGS-----EGSAISGVASSTATL- 545

```


460 EVKTDASLPSESEK-----PEKEISKPVVTVSSRRNGSVRKQTSVAVSPPOEDRENDLV 511
461 --MOLSKIAMTSLPMSMALLTNHFKSTGSPFLPCARALGAS-----PSETSKLOOLV 594
512 EHNINIAKLVRSSA-----SHESQAEVSLAQMERICIDSKSWEDLIEIDKTSKLOLV 566
595 EKDIDROGAVAVTSAASGAPPTTSAAPSSSASSGPNOCVLCRLVSCPRALRLHYGQHGGE 654
567 DNIENK-----LTPDNOCIFCOKVWSCSSLSQMHTRTHGE 602
655 RPRCKYKGRAFSTRCGNLRAHPVGHKASPARARONCPCIOCKFTNVTLOOHYMHU- 712
603 RPRCKTICGRAPFATKGNLKAHMSIHKIKPRMNSQFCPCVCHQAFNSGIILOOHIRIHTMD 662
713 ---GGQ-IPNGCTALPEGGAOENGSS-----EQSTVSGAGSEFPQ 748
663 DSGSGGGVPAANPGFAERLGIQDQNSNKSLGTSDTLDFSTTISDHSQGRSESSQCGDDE 722
749 QOSQO-----PSPEPELSEEEDEDEEDVDYDEDS-----LAGRGSSES 788
723 FMTWDSSTDSDSDNSAATATAPPHLEREREREKERRRIPNDCCSDERSHNPDLTGGRSES 782
789 GGEKAISVRCGSEEA-----SGAEEVGT 812
783 GEMPAAMDLSFSSNSGRTFATGLANGATGGSGNGGLPMLGMPMPNLLMAAAREEMHA 842
813 VAAA-----AT----- 818
843 LGHNAKAPFLPFGPLGFWGLHPRPVNCNLCFPMPLSLALLESHLQSEHAKEPATGHQNR 902
819 -----AGKEMDSNEKTQOSSLP-----PPPP-DSLD 845
903 PHCSDAGSPYCAKTLNPLNPLFAKKPPSSSSSSGEEKLPESINPPPAENPPATPTKEDPD 962
846 QOPQM-EOGSSGCVLGKKEGKGKPERSSSPASALTPEGEATSVTL-----VBEISLQEA- 897
963 QEOLMVEGAS-----AGESSGCGATSNYPQEGADAEGLMKMOJLHAHFPAASPLDFOAL 1018
898 MKREGGESS-----SRKACEVCGQAFPSQALAEHOKTHKEGRLFCVFCRQGFLEA 951
1019 MSAGPPTSLDPPVNNKHCHVCRRNFSSSSALQIHMTHRGDKP-FQCNVCOKAFATTG 1077
952 TLKKHMLAHQVOPFAPHPONIAALSLVPG 983
1078 NLKVHM-GTHMNTNPISRRGRKRSLELPMRPG 1108
RESULT 13
09YKX3 PRELIMINARY: PRT: 1267 AA.
09YKX3; AC
01-MAY-2000 (TREMBLrel. 13, Created)
01-MAY-2000 (TREMBLrel. 13, Last sequence update)
01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE
SALR protein.
GN
SALR OR CG4881.
OS
Drosophila melanogaster (Fruit fly)
OC
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC
Ephydroidea; Drosophilidae; Drosophila.
OX
NCBI_TaxID=7227;
[1]
SEQUENCE FROM N.A.
RP
STRAIN-BERKELEY;
RC
MEDLINE=20196006; Pubmed=10731132;
RA
Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA
Amenaides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA
George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA
Stuton G.G., Mortan J.R., Vandel M.D., Zhang O., Chen L.X.,
RA
Brandon R.C., Rogers Y.-H.C., Blazet J.G., Champs M., Pfeiffer B.D.,
RA
Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA
Abrill J.F., Abdanyani A., An H.-U., Andrews-Pfannkuch C., Baldwin D.,
RA
Bailew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,

RA
Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA
Borkova D., Botchan M.R., Bouck J., Brokstein P., Bottler P.,
RA
Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA
Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA
de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA
Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA
Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA
Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA
Glodex A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA
Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA
Hostin D., Houston K.A., Howland T.J., Wei M.-H., Idegawa C.,
RA
Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA
Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA
Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA
Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA
Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA
Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA
Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclob J.M.,
RA
Palazzo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA
Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA
Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA
Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA
Sylvitskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA
Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
RA
Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA
Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA
Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA
Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
RT
"The genome sequence of Drosophila melanogaster."
RL
Science 287:2185-2195(2000).
DR
EMBL: AE003632; AAF53096.1; -.
DR
HSSP: P15822; 1BRO.
DR
FlyBase: FBgn0000287; salr.
DR
InterPro: IPR003006; Ig_MHC.
DR
InterPro: IPR008822; ZnF_C2H2.
DR
Pfam: PF00096; zf-C2H2; 8.
DR
SMART: SM00355; ZnF_C2H2; 8.
DR
PROSITE: PS00290; IG_MHC; UNKNOWN_1.
DR
PROSITE: PS00028; ZINC_FINGER_C2H2_1; 8.
DR
PROSITE: PS00577; ZINC_FINGER_C2H2_2; 7.
KW
DNA-binding; Metal-binding; Zinc-finger.
SQ
SEQUENCE 1267 AA: 139540 MW: 45101DBICOP47829 CRC64;
Query Match 13.8%; Score 726; DB 5; Length 1267;
Best Local Similarity 23.3%; Pred. No. 2.4e-34;
Matches 276; Conservative 141; Mismatches 390; Indels 378; Gaps 40;
QY 67 QENPNNSASSEPRPEGHNNPQVMDTEHSNPPDGSQSVPTDPTWGPBERGESSGHF--- 123
DB 38 RERDRDTLLAKEILDADSDNNNGTEPOMEAEAVPESPTERTAEERGEDEPENSNALDLS 97
QY 124 IYAATGTAGGGGGLI-----LASPKIGATPLPESTPAAPRPPPPRPGVSGH 174
DB 98 LISSGRESIPGSGHVSLEALQHTKVAAPALATMAGNHQSA----- 140
QY 175 LNIPLILEELVLOQROIHQWMTQEOICROY-----LLLG--SLQGTGAPASPSELPGT 227
DB 141 -DLAWQSTIFNVQRIQLMQQLQIHLQHSQLRKABAAALGRHSHSDEEEERPERPK 199
QY 228 GIASSTKPLPLPLPSPIKPVQTSKTLASSSSSSSSSGATTPQAQFNLHLYPLDSQHFSS- 286
DB 200 QPTNGLKEELELEQGRPESEDOES--RRENSKTKDRKGTDRKA-----EREQSQSMCD 252
QY 287 -----AGVGRSHKPTAPSP-----PALPGSDQTLASPLHAFPSYTGILAAQCLGARG 336
DB 253 ISSSLASITTNMDPPRAPRPNPCLEMLQRIKEVYLD-----ASQSLHAAQ- 299
QY 337 LEATAPGLKLPKNGSELSYGVWGLPEKRGKRCFCAKYGSDALQIHRLSTGE 396
DB 300 MQEVEYSYASKAEGASRGHIF-----KHRCCKCGKIFGYSALQIHRLSHTE 346
QY 397 RPRKCNVCGNRTFRGNLKVPHNRHREKYPHYOMNHPRENHDYVITSSG----- 447


```
FT  NON_TER      1      1
SQ  NON_TER      327    327
    SEQUENCE      327 AA: 35908 MW: 1623D9E63F7E96689 CRC64:
Query Match
Best Local Similarity 12.3%; Score 648.5; DB 13; Length 327;
Matches 145; Conservative 46; Mismatches 81; Indels 89; Gaps 10;

OY  404 CGNRFTTGRGNLKVHFRHRRKRYPHVQNNPHRVPPEHLVDYITSSGLPYGMSVPEKAE643
    1 CGNRFTTGRGNLKVHFRHRRKRYPHVQNNPHRVPPEHLVDYITSSGLPYGMSVPEKAE643
    1 CGNRFTTGRGNLKVHFRHRRKRYPHVQNNPHRVPPEHLVDYITSSGLPYGMSVPEKAE643
OY  464 ATPGGGVERKPL---VASTATLSTESLTLLSTAGTATA-----PGLPAF---- 506
    60 -----WLDKPLIPTITPTISVALQLPPTIPSTIGSGDSTSPSLKRSPPRPSPSSECTS 114
    507 ---NKFVLKAV-----EPKKNKADENTPPSEGSAT-----SGVAESSTATLMQL- 548
    115 LSPNHLITETSLAQISSPPQPNLASNTPPVLPKPEALHLPTNSTRRGETSISTASISQVI 174
OY  549 -----SKMTSLPSPALLTNHFKSTGSPFLPLCARALGASPSSETS 588
    175 STTIVTTTCSTRTQLTDPVNVSSAASHPSLSQISSNQFNP--KPEGGGLDSM--QTSETS 230
OY  589 KLOQLVERIDROGAVAVTSAASGAPPTSAPAPSSASGPNOCVICLRVLSCPRALRLHY 648
    231 KLOQLVERIDRK-----MTDPNOCVICLRVLSCOSALMKMHY 266
OY  649 GONGERPEPKCKVCGRAFSTRGNLRAHFVGHKASPARAQNSCPICQKKFTNAVTLQOHV 708
    267 RHHTGERPEPKCKICGRAFTTKGNLTKHFVGRSKRPLRVQHSCPIQCKKFTNAVVLQOH 326
OY  709 R 709
    327 R 327

RESULT 16
OY  09DGH2 PRELIMINARY; PRT; 288 AA.
AC  09DGH2:
DT  01-MAR-2001 (Tremblrel. 16, Created)
DT  01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT  01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE  Putative spalt protein (Fragment).
GN  SpALB.
OS  Brechyanlo rerio (Zebrafish) (Zebra danio).
OC  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC  Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC  Cyprinidae; Danio.
OX  NCBI_Taxid=7955;
RN  [1]
RP  SEQUENCE FROM N.A.
RA  Camp E.M., Jardelli M.T.;
RT  "PCR amplification of spalt exonlc sequences from various
    vertebrates."
RL  Submitted (Aug-2000) to the EMBL/GenBank/DBJ databases.
RL  EMBL; AJ293863; CAC05356.1; -
DR  InterPro; IPR000345; CytC_heme_bind.
DR  Pfam; PF00096; zf-C2H2; 2; Znf_C2H2.
DR  SMART; SM00355; Znf_C2H2; 4.
DR  PROSITE; PS00190; CITOCHROME_C; UNKNOWN.1.
DR  PROSITE; PS00028; ZINC_FINGER_C2H2_1; 1.
DR  PROSITE; PS50157; ZINC_FINGER_C2H2_2; 3.
KW  DNA-binding; Metal-binding; Zinc-finger.
FT  NON_TER      1
FT  NON_TER      288
SQ  SEQUENCE      288 AA: 32261 MW: A07A1F2BBD6A6F42 CRC64:
Query Match
Best Local Similarity 12.0%; Score 631; DB 13; Length 288;
Matches 143; Conservative 40; Mismatches 67; Indels 92; Gaps 9;
```

```
OY  404 CGNRFTTGRGNLKVHFRHRRKRYPHVQNNPHRVPPEHLVDYITSSGLPYGMSVPEKAE643
    1 CGNRFTTGRGNLKVHFRHRRKRYPHVQNNPHRVPPEHLVDYITSSGLPYGMSVPEKAE643
    1 CGNRFTTGRGNLKVHFRHRRKRYPHVQNNPHRVPPEHLVDYITSSGLPYGMSVPEKAE643
OY  464 ATPGGGVERKPLVASTATLSTESLTLLSTAGTATAPGLPAPNKFVKKAVPPKNKADE 523
    56 -PPLSWLDSKPLVGNLSIGFMSSSL-----PGLPV----IIRKEEGVSTIKP 98
OY  524 NPPGSE-GSAISGVAES-----STATLMQLSK----- 550
    99 HSPVSELTGRINGHOGEGFVCSPLIISNEKFOEVNQRLSVSTLRSREGSREDIAINTS 158
OY  551 ----LMTSLPSPALLTNHFKSTGSPFLPLCARALGASPSSETS KLOQLVERIDROGAVAVT 606
    159 VNTGLTKIKSEGLEAKFL--LGSLPNPLGA-----SETSKLEQLENDIRK----- 203
OY  607 SAASGAPPTSAPAPSSASGPNOCVICLRVLSCPRALRLHYGONGERPEPKCKVCGRAF 666
    204 -----YDPNKGICICLRVLSCOSALRMHFRTHGERPEPKCKVCGRAF 245
OY  667 STRGNLRAHFVGHKASPARAQNSCPICQKKFTNAVTLQOHV 708
    246 TTKGNLTKHYSIRHSMPLRLQNSCPICQKKFTNAVVLQOH 287

RESULT 17
OY  09PVN5 PRELIMINARY; PRT; 549 AA.
AC  09PVN5:
DT  01-MAY-2000 (Tremblrel. 13, Created)
DT  01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT  01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE  Xsal-3', long form (Fragment).
GN  Xsal-3'.
OS  xenopus laevis (African clawed frog).
OC  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC  Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
OC  Xenopodinae; Xenopus.
OX  NCBI_Taxid=8335;
RN  [1]
RP  SEQUENCE FROM N.A.
RA  MEDLINE-99458636; Pubmed-10527856;
RA  Onuma Y., Nishinakamura R., Takahashi S., Yokota T., Asashima M.;
RT  "Molecular cloning of a novel Xenopus spalt gene (Xsal-3).";
RL  Biochem. Biophys. Res. Commun. 264:151-156(1999).
DR  EMBL; AB030826; BAA85902.1; -.
DR  HSSP; P15822; 1BBO.
DR  InterPro; IPR000823; Znf_C2H2.
DR  Pfam; PF00096; zf-C2H2; 5.
DR  ProDom; PD000003; Znf_C2H2; 1.
DR  SMART; SM00355; Znf_C2H2; 5.
DR  PROSITE; PS00028; ZINC_FINGER_C2H2_1; 5.
DR  PROSITE; PS50157; ZINC_FINGER_C2H2_2; 5.
KW  DNA-binding; Metal-binding; Zinc-finger.
FT  NON_TER      1
SQ  SEQUENCE      549 AA: 59348 MW: ECBBA478D0F088A46 CRC64:
Query Match
Best Local Similarity 9.7%; Score 513; DB 13; Length 549;
Matches 134; Conservative 30.5%; Pred. No. 2.7e-22;
Mismatches 148; Indels 98; Gaps 13;

OY  585 SETSKLOQLVERIDROGAVAVTSAASGAPPTSAPAPSSASGPNOCVICLRVLSCPRAL 644
    5 SETSKLOQLVERIDRK-----SETNCLICLRVLSCPSSL 40
OY  645 RLHYGONGERPEPKCKVCGRAFSTRGNLRAHFVGHKASPARAQNSCPICQKKFTNAVTL 704
    41 KMHTRHHTGERPEPKCKICGRAFTSKNLTGVRANRPLKQHSPIQCKKFTNAVVL 100
OY  705 QOHVRHMLAGQIPNGATALPEGGAOENGSEOSTVSGA--GSFPQOQ----- 750
    101 QOHIRHMGCKIRP-----TPVSEASDDIDSMMDEKNGELNLSFTDENLDDIMDEDELA 156
```



```
250 HHPVLAQOGS-----GSSPKATDIPASAPPPVAGVPFFKQSPGHQSPLASPRVVPQCP 304
243 IK-----PVQSKTILASSSSSSSSGAF-----TPKAFPHLHYHLSQHP-- 285
305 LKEEDDESPVDS---SPQSPSSSGAEADSDNSPASS-----SSRPLKV 351
286 -----SAGVGR-----SHKPTAP-----SPALP----- 305
352 RIKTIKTSQGNIRITYQVSDPPRALBAGAFLEASLILKISPATPTSEGPVVSQVL 411
306 -----GSTDOLA-----SPHLAPSTTGLLAQCL 331
412 GDGTRIKGTVLPAATIONASTAMLAASVARKAVLPGGTATSPKMAIKVLIHV-PCAL 470
332 GAARGLEATAPG-----LLPKNGSGEL----- 355
471 PKADGAGAGTGGOKYNGASVVMQSPKATGPSTGGTIVISHTQSSLYEAFNKILSKN 530
356 ---SYGEVNGP-----LEKPGRHKRCFAKVFSGDSALOIHLSHTGERPRYKCNVGNR 407
531 LLPATRPNISPPRAEAGALAPPTGYRCLECGDAFLEKSLARHYDRSMRELVTCNHQARR 590
408 --FTTRGNLKVFHNRREK-----YPHVQNPHPVPEHLDYVITSSGLPYGMSVPREKAE 460
591 LVFFPKSCILLHAREHKDGLVMQCSHLMPRVALDQWVGQPDITPLP--VAVVP--VS 646
461 EEAATPGGVERKPIVASTALSTALSTESLT--LSTAGTATAPRLPAFNKVLMAKAVEKN 519
647 GPLALPALKBGSAITSSAITTYAAEAVLP--LSTE---PPAAPTASATCFCECKE 701
520 KADENTPPGSEGAISGVAESSTATLMQSLMTSLPSSWALLTNHFKSTSPFLPLCARA 579
702 QCRDAGMAAHFQQLGPRAPGATSNVCTCPMM--LPNCSFSAHQRMHKNRPHVCEPC 759
580 LGASSETKLOLVEKIDQAGAVAVTSAAGAPTTSAAPSSASAGNOCVYICLRVLS 639
760 GG--NFIQANFQTHLEACLHVSRRVG-----YRCPSCSYVFG 795
640 CPRALRLHNGOGERPFCQKVCGRAFSTRGNLRAHFVGHKASPARAONS-----CPIC 694
796 GVNSTKSHQTSICEVFHKPCICPMAFKSGPSAHNL--YSHQPSQTQOAKLITKCAAC 853
695 QKFTNAVTLQOQVHRMHLGQOIPNGGTALEPGGGAQENGSEOSTVSGAGSFPOQSQOP 754
854 DTFVHKRPLLSHFQDHL-----LP-----QGVSVFKCPSCPLFAQKR 892
755 SPEELSEEEDEDEEDVTDSDSLAGRSESGEKAISVAGDSEASGAEEVGTVA 814
893 TMLEHLKNTHQSGLEE-----TAGKA--GGALLTPKTEPEELVVSQ-----G 934
815 AAATAGKEMDSNEKTTQOSSLPPPPPSLDOPOMEGSSGVLGCKEKGK----- 866
935 GAAPRTSESSSSSEEEVPSPPRP--AKRPR-RELQSKILKG--GGPGGWTGGLC 989
867 ---DERSSSPASALTPREGATSVTLVEELSLQEAARKKPGSSSSRKACEVCGQAPSOA 922
990 HSMFPERDEYVAH-----MKKEHGKSVKKFPCRLCERSFCSAP 1027
923 ALEENOK-THPKEGRLTYVFCROG---FLERATLKKNHLLAH 961
1028 SLRRHVRVNHGKIKRVPKRYCPEGKRTSSRLLEKHQVHRH 1070

RESULT 21
61479
061479 PRELIMINARY; PRT; 2282 AA.
01-NOV-1996 (Tremblrel. 01, Created)
01-NOV-1996 (Tremblrel. 01, Last sequence update)
01-MAR-2002 (Tremblrel. 20, Last annotation update)
DNA binding protein RC.
KRC OR RC.
```

```
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/C; TISSUE=THYMOCYTE;
RX MEDLINE=97001141; PubMed=8812474;
RA Wu L.C., Liu Y., Strandmann J., Mak C.H., Lee B., Li Z., Yu C.Y.,
RT "The mouse DNA binding protein Rc for the kappa B motif of
RT transcription and for the V(D)J recombination signal sequences
RT contains composite DNA-protein interaction domains and belongs to a
RT new family of large transcriptional proteins."
RL Genomics 35:415-424(1996).
CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
DR EMBL; L46815; AAA40884.1; -.
DR HSP; P13822; IBOO.
DR MGD; MGI:106589; Krc.
DR InterPro; IPR000822; ZnF_C2H2.
DR Pfam; PF00096; zf-C2H2; 5.
DR PRINTS; PR00048; ZINC_FINGER.
DR SMART; SM00355; ZnF_C2H2; 5.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 4.
DR PROSITE; PS01517; ZINC_FINGER_C2H2_2; 5.
KW DNA-binding; Metal-binding; Nuclear protein; Zinc-finger.
SQ SEQUENCE 2282 AA; 246919 MW; E4404B8C9F7B2C CRC64;
```

Query Match 6.6%; Score 346.5; DB 11; Length 2282;
Best Local Similarity 23.0%; Pred. No. 6.1e-12;
Matches 224; Conservative 114; Mismatches 336; Indels 299; Gaps 44;

```
QY 203 RQVLLGSLGQTVGAPASPELPGTGA---SSTKPL--PLFSPKRPVQTSKTLASS 255
DQ 18 RKRLTKGEALIQTSVSSAP--YPGSGTTAPSESATQELLAQPPSGP-----SQ 64
QY 256 SSSSSSSSGAETPK-QAFTHLHYHLSQHPSPAGVGRSHKPTPA-----PSPALPGST 308
DQ 65 EKTGGQOKPARRPSIEASVHI-----QLP-----QHPLTPAFMSPGKPHLLEGST 111
QY 309 DOLI---ASPHLAF-----PSTTGLLAQCL-GAARGLEATAPRLLPKNG 351
DQ 112 WQLVDMRMRPGSGSEFAPGSHPOSQLPSHNSILPPEELPGLPKVPVPPSOVSLRP-- 168
QY 352 SGELSYGEVNGPDEKPGGHNKRCFAKVFSGDSALOIHLSHTGERPRYKCNVGNFTTR 411
DQ 169 -AEENHKKERK-QKR-GKTYICQYCSRCAKPSVLQKIIRSHGTGRPYCGPGCFEKK 225
QY 412 GNLKVHFRHREKYPHVQNPHPVPEHLDYVITSSGLPYGMS---VPREKAE----- 461
DQ 226 SNL-----YKHKRSHAIRIKAGLAGSSSEMYPGLEMERIPGEFEFEPTGEEST 275
QY 462 -----EAATPGGGERKP-----LVASTALSTESLTILSTLSGTA---TAPGL 503
DQ 276 DSEETGAAGSPSTDVLPKPHLLSSLSYSGSHGSOERCSLSQSGTSPSLIEDPAPFA 335
QY 504 PAFNKFVL-----MKAVEPRNKADENT--PGSEGAISGVAESSTATLM 546
DQ 336 EASSEHPLSHKPEDHTTIQKIALRLSEKKKLIIEQTLPSGSKSTEGSYRSASAQ 395
QY 547 QLSKMLTSLPSMALL-----TNHFKSTGSPF-LPLICAR---ALGASPSSETSKL 590
DQ 396 QVSPNTNAKSYAEILFGKCGRIGORTSMLASTQPLPLPSSEDKPSLVLPSVPTQVY 455
QY 591 QQLVEKIDROGAVAVTS-----AASGAPTSAPAPSS----- 623
DQ 456 EHITKLITINEAVVDISEIDSVKPRRSSLTRRSSVESPKSSLYRDSLSSHGETKQEOSL 515
QY 624 -----ASSGNOCVICTRLVLSCPRALRLHYGOHGERPCKKCGRAFSRGNLRAPF 676
DQ 516 LSLQHPSSSTHP---VPLLRSHMPSA-----ACTISTHHTTRGST--SF 556
QY 677 VGHKASPARAONSCPICQKFTNAVTLQQHVNR--HLGQIIPNGGTALEPGGGAQEN 733
```


Db 557 DDHVADEVPSPNT-PV-----FTS-----HPRMLKRHAALPLG----- 592
QY 734 GSEOSTVSGASSFPQOOSQSP-EBELSEEEDEDEEDVD-----ED 779
Db 593 -----EYSSEEPGPPSKDPSTSKPDEDEPEKEDLTKTKKGKTKKANEC 638
QY 780 SLARG-----SESGEKAISVRGDSSEASGAEVGTVAATAAGK--E 822
Db 639 TIGCARVKKRNDYEAKKKYCELOITKAHSVGAHEVEKTAEPFPMQMMHYKIGATLE 698
QY 823 MDSNEKTTQOSSLPPPPPSLDQOPMEQSSGVLGKEEGKPERSSPASPALTPEGE 882
Db 699 LTPLRKRREKSLGDEEPPAFACPGPSEFAHNRLGSTK---SPAFAKSAAPSL-EDPR 754
QY 883 ATSYLVLELSIQEAMRKPERGSSSKACVCGAFPSQALBEHQKHKPKGPLETCVF 942
Db 755 ASSPQLPSQ-----ELGQNGQRRG-EQCPKFTVYIQTSSFEKEDPPDP----- 798
QY 943 CROGFLEBATELKKHMLAHQVQFAPRG-----PQNTAALSLVPGCSPTS 990
Db 799 -----SGLEEDKPRPAPSSPPRAPHGRSAHSLQRLVAKQNPQLVTEEPDRD 850
QY 991 TGLSPFRKDPPT 1003
Db 851 TEPEPPPEPEKT 863
RESULT 22
Q9NS43 PRELIMINARY; PRT; 751 AA.
ID Q9NS43 AC Q9NS43;
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Hypothetical zinc finger-1-like protein (fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP
RA Kodojann V., Ge Y., Severin J., Krummel G.K., Grable L.,
RA Kvikstad E., Gordon L., Brower M., Olsen A.S., Smith L.M.;
RA "Sequence analysis of a 1mb region in 19q13.2 containing a zinc finger
gene cluster";
RT Submitted (Jul-2000) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
DR EMBL; AC074331; AAF88107.1; -.
DR HSSP; P08045; 1ZNF.
DR InterPro; IPR001909; KRAB.
DR InterPro; IPR000822; ZnF_C2H2.
DR Pfam; PF01352; KRAB; 1.
DR Pfam; PF00096; ZF-C2H2; 18.
DR PRINTS; PR00048; ZINCFINGER.
DR SMART; SM00349; KRAB; 1.
DR SMART; SM00355; ZnF_C2H2; 18.
DR PROSITE; PS00805; KRAB; 1.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 18.
DR PROSITE; PS0157; ZINC_FINGER_C2H2_2; 17.
DR DNA-binding; Metal-binding; Nuclear protein; Zinc-finger.
KW NON_TER
FT SEQUENCE 751 AA; 86207 MW; 96583BA581c87fc CRC64;
SQ
Query Match 6.58; Score 342; DB 4; Length 751;
Best Local Similarity 21.18; Pred. No. 4e-12;
Matches 187; Conservative 81; Mismatches 335; Indels 282; Gaps 29;
QY 170 VSGSHNIPILIE--LRLVQQRQHOMQMTQICROV-----LLG----- 209
Db 42 VSGKHQNMKMTLQKALKYLSNQLSCWQIKQWASSELTRCLQSGSSQLQDGIQVSEN 101
QY 210 -----SLGQTVGAPASPSSELPGTGTASTKPL----- 237

Db 102 ENNINPKGDSSIIYIENOEFPFRMTOHSCGNTY---LSEQIQSRKQIDVKNINQIHEH 158
QY 238 -----PLSPFKPVQTSKTLASSSSSSSSGAETPKQAFHLYHPLGSO-HPFASGVG 291
Db 159 FMKASPFHNIKTDTPRKPGNEYGKIIIDGSNOKL-----PLGEKRPFGCGG 210
QY 292 RSHKTPAPSPALPGSTDOLIASPHLAFPTTGLLAQCLGARNLEATAPGLKPK-- 349
Db 211 FSY-----SPRLPLPHN--VHTGKCFQSOSHLRTHQIHGKELNRCHESGDCEPKSS 262
QY 350 -----NSGSELSTGYEVMGLEKPGGNHKCFKCAKVFSGDSALQIHLNHSHTBERPKCV 403
Db 263 FHSYSNHTGKESY-----KCDSCGKGFSSSTGLIHYRHTGEKPKCEE 308
QY 404 CGNRFTFRGLKVFHNRHREKYPHVQNNPHVPEHLDVITSSGLPYGMSV----- 454
Db 309 CCKCFQSQSNFOCHQVHTTEK-----YKCECGGFGFQSVNLRVHQRVH 354
QY 455 ---PREKAEENATPGGVERKPLVASTTALATSESLTLLSTAGTANAPGLPAFNKPYL 511
Db 355 RGEKPYKCE-----CGKGFQAAPHI 377
QY 512 MKAV---EPKNKADENTPPGSEGAISGVAESSTATLMQSLKMTSLPSMAL----- 560
Db 378 HQRVHTGEKPYKCD-----YCGKGFSSNPLICHRVHTGEKPYKCEACGKFT 426
QY 561 ----LTNHF-K-STGSFPLPCARALGASPSFETSKLOQLVEKIDROGAVAVTSAAGAPTT 615
Db 427 RNTDLIHFRVHTGEKPYK---CKEKGKFSQASNLQVQNVHTGKFRFCPTCGKFSOS 484
QY 616 SAPAPSSASSG---PNQVCICLRVLSCPRALRLAHGQNGERPFKCVCGRAFSTGNLR 673
Db 485 SKLQTHQVHTGKPYKCDVCGKDFSYSNLKLHQVHTGKPYKCECGKGFSSMSNLH 544
QY 674 AHFVHGKASPAARQNSCPICQKFTNAVTLOQVHBMHGGQIPNGTALPEGGAQEN 733
Db 545 AHQRVHTGEKPYK---CEQDKSFQALIDFRVHQVHTGKPYKGCVC---GKGSQSS 597
QY 734 GSEOSTVSGASSFPQOOSQSP-EBELSEEEDEDEEDVDDEDLSLAGSGSEGEKA 793
Db 598 GLQSHQVHTGEKPY-----KQDVCGKGFGRYSQFI 628
QY 794 ISVRGDSSEASGAEEVGTVAATAAGKEMDSNEKTTQOSSLPPPPPSLDQOPMEQ 853
Db 629 YHQRGHTGEKPYKCEEG---KGFGRSL-----NLRHQVHTG 664
QY 854 SSGVLGKEEGKPERSSPASPALTPEEATSVTLVEELSLQEAMRKPESSSRK--AC 911
Db 665 EKPHI--CEECGK-----AFSLPSNLVNLGVHTREKLPKC 698
QY 912 EYVCGAAPPQALBEHQKHKPEGFLTCVFCROGFLEBATELKKH 956
Db 699 EECGKGFSGARLEAHQVHTGEKPYKCDIDCKDBRHHSRLTYH 742
RESULT 23
Q9UEG4 PRELIMINARY; PRT; 927 AA.
ID Q9UEG4 AC Q9UEG4;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)
DE KIAA0326 protein (Fragment).
GN KIAA0326.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-BRAIN;
RX MEDLINE=97349984; PubMed=9205841;
Nagase T., Ishikawa K., Nakajima D., Ohira M., Seki N., Miyajima N.,

566 HMDENL.FVCSDCGKFLAEHLEQHR-----VHERGKTPARRAQ----- 605

Db 158 HÖRGHGERPYACADCGKSFADPS---VFRKRR--THAGLRP-----YSCERCCK 203

309	HURSHUENKELACNVCNKKK	IIKGINLVAHNE	HNKHNKNI	PHVQNNPPIV	VEEHLDI	VIIISGOL	440
150							
100							
50							
0							

OY	449	PGMSVPRPEAKEEEDATPGGVEKKPLVAST--TALSTESTLTLSTAGTATAPLP	506
Db	204	AVA-ELKDLRNHRSHST---GERPLCSGCKSFSRSSSLTCORHIAAOKPYKSPAC	257
OY	507	NK-FVLMKAVEPKNKADENTPRGSEGAISGVABESTATLMQSLKMTSLSPWALLTNF	565
Db	258	GKGFTOLSSVQSHERTHSGKP-----	279
OY	566	KSTGSEPLPLCARALGASPESTSKLOLVKID-----ROGAVALVTSASGAPT	614
Db	280	-----FLPCRCGR-MFSDPSSFRHGOAHNEGVKPYHCCKCKDFRQ-----PA	321
OY	615	TSAPASSSSASSPNOCVITLVLSCSRALYLYHGNGGERRPKCKVCGRASTKGNLA	674
Db	322	DLAMHRVHTGDRPFKLCLODKTFEVAWMDLKRHALVHSGORFRCCEGCRAFERAFLTK	381
OY	675	HEVGNKASPARAARONSCPRCKKPTANVTLOONRYMNGGQ---PRNG-----GTALPRG	726
Db	382	HEVNH-----SGERRPHCNACGKSTVYSSLSLKHERTHSSSEAGVPPRADELVGLALPV-	436
OY	727	GGAADENGSEOSTVSGAGSFPPQOOSQORSPREELSEEEEDDEEDVTDSDSLGRGS	786
Db	437	-VWAGESSAAPAGACIGDPPA-----GILGLRP	464
OY	787	ESGGEKA-----ISRGSEEAASGADEEYGYAAATATAGKEMNSNETTQOSSLPRP	839
Db	465	ESGGVATMOQVVGVTVEHVECDAGVREARPLLEGAGE-BADEK-----	512
OY	840	PRDRLDOPRMEGSSGVLGKGEGKRRPERSSPASALTPEGEATSVTLVEELSLQEAR	899
Db	513	PRQFVGR-ECKEFTSTLTLRNHRSHRLRPP-----	545
OY	900	KEBGESSRKACSEVCGAARPSOAALEENOKTHPKERGLTCVFCROGFLERATLKKHML	959
Db	546	-----CTCCGKSISDRAGLRKHSRHSVSR-YTCRNCKAFASLDRKHE--	591
OY	960	ANHOVRAPRPHGRONIALSLVRECCSS 967	
Db	592	RTHVRPMGTRPLRLPVALLGMPREGPA 619	
RESULT	25		
O960L6			
AC	O960L6:	PRELIMINARY: PRT: 736 AA.	
DT	01-DEC-2001 (TREMBLrel. 19, Created)		
DT	01-DEC-2001 (TREMBLrel. 19, Last sequence update)		
DT	01-JUN-2002 (TREMBLrel. 21, Last annotation update)		
DE	LD462339.		
GN	CG12299.		
OS	Drosophila melanogaster (Fruit fly).		
OC	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;		
OC	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;		
OC	Ephydroidea; Drosophilidae; Prosochilla.		
OX	NCBI_Taxid=7227;		
RN	111		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=BERKELEY;		
RA	Stapleton M., Brokstein P., Hong L., Aabayani A., Carlson J.,		
RA	Champe M., Chavez C., Dorsett V., Farfan D., Fiske E., George R.,		
RA	Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,		
RA	Nunoo J., Pacleb J., Pargass V., Park S., Phouanavong S., Wan K.,		
RA	Yu C., Lewis S.E., Rubin G.M., Celniker S.,		
RL	Submitted (Aug-2001) to the EMBL/GenBank/DBJ databases.		
DR	EMBL: AY051994; AAK93418.1; -		
DR	FLYbase: FBgn0032295; CG12299.		
DR	InterPro: IPR000822; Znf.C2H2.		
DR	Pfam: PF00096; zf-C2H2; 10.		
DR	PROSITE: PS00028; ZINC_FINGER_C2H2_1; UNKNOWN_10.		
DR	PROSITE: PS00157; ZINC_FINGER_C2H2_2; 10.		
DR	DNA-binding: zinc-finger		
SO	SEQUENCE 736 AA; 81670 MW; 0B1C198D547B4C2C CRC64;		

Query Match	6.38	Score 333.5	DB 5	Length 736
Best Local Similarity	21.0%	Pred. No. 1.2e-11		
Matches 179	Conservative 82	Mismatches 279	Indels 311	Gaps 31
QY 142	SPRLGATPLRPPESTRPAPRRPPRRPPRGVSGSHNT-----PLL-EELRVLQDQOIH	193		
DB 17	TPPLQOHTHLLPQQAAPPNPQQPQQPQPPDLTRHCMCAEEFVHPLALYQHNHTLHPHRRG	76		
QY 194	OMO-----MTEQICQVLLLSLGGTVGAPASPSLPGTSTASSTKPLLP	238		
DB 77	NGQOQESPDGEDSDYSWIFEPVC-----ELAEQSSSDSGSASGSSSSSD-----	125		
QY 239	LESPIKPVQTSKTLASSSSSSSSSSSGAETPKQAFHLYHPLGSOHPS--AGVGRSH--	294		
DB 126	-----DDDDDDDDSSSSSSSSSSSSSSSVPTTNSMTQSQESVQPLHGLVAGGVNEFQ	183		
QY 295	-----KTPAPSPALPGSTDLI-----ASPHLAPSTTGLLAQCSGAA	334		
DB 184	LQMTDPRESTSIEMVQAPVSVTP-----LQQLLPAPVSPGIGLQSTP-----IKRR	231		
QY 335	RGLEATASPGILKPK--NGSG-----ELSYGEVGPRLKPCGRH-----KCRFCAKV	379		
DB 232	RGGRSINIGAVMDPALNGOKCFQCTHCEASPRNA--GDSKHYRSHITTKPFQCSIOCT	290		
QY 380	FGSDSALQILHRSHTGERPYKCNVGNRETTKGNLKVHNHREKYPH--VQNNPVPYE	437		
DB 291	FTIHGSLNTHIRHISGERPKYCELCPKAFPTQSSSLVHMRSHSVRKHCQVQCKG----	346		
QY 438	HDYVITSSGLPYGMSVSPREKAEAEALPFGGVEKRPVLASTALSTESITLLSTAGT	497		
DB 347	-----FINYSYL-----LHQKHTIAPTETF-----	367		
QY 498	ATAPGLPAFKFVLMKAVERPKKADENTPPGSEBSGASISGVAESTATLMQLSKMTSLPS	557		
DB 368	-----ICPECEEFKAE-----	379		
QY 558	WALLTNHFSTGS--FPLPLCARALGASPSSETSKLQOLVEKIDROGAVAVTSAAGAPT	614		
DB 380	-ALDEHMRNHTQELVYQCALCREAFRAS-----SELVQHKKNH-----MGKRF	423		
QY 615	TSAPADPSSASSG-----PNOVCYCLRVLSCPRALRYLHGNGGERPFKCV	661		
DB 424	TCSLCDRSEFTQSGSLNTHMRITHTGCKPFOCKLCKDCFTQASSLSLVNKKIHAGEKRP	483		
QY 662	CGSAFSTRGNLRAHFVGHKASPARAQS-----CPSCQKFTNVTLOQH	707		
DB 484	CGKSTSQQAVLTKHQAOMKASASASTSPGLLVAKOPHETLVCIVGSLHADATLASH	543		
QY 708	VR-----MHNGCOIPNGGTALPEGGGAQENGS-----EQ	737		
DB 544	VHSQHALLDTMKQSGMTPARAGATPDVKCSABEOQAYVERVOCVLODMNHQOHOHQO	603		
QY 738	STVSGAGSPPOQSQO-----PSPEBELSEEEED	768		
DB 604	QPPQOQOQHPPQOQOQHLLQOQPHQWQLPQOPKPLPAMDSTGEDEEERPDAAERPPDEE	663		
QY 769	EEEE-----EDVYDEDSLAGRSBSGG--EKAIYSRGD-----SEASG--	805		
DB 664	EEPEPAEVKTEVLVAEDALINPGYPIPLGLEEOIILDSMWYUEDFGMDVGCQGEVEFDF	723		
QY 806	--AEEEGTVA 814			
DB 724	VVNEEEVYTDA 734			
RESULT 26				
ID 015090	PRELIMINARY;	PRT;	1300 AA.	
AC 015090:				
DT 01-VAN-1998 (Tremblurel. 05, Created)				
DT 01-VAN-1998 (Tremblurel. 05, Last sequence update)				
DT 01-JUN-2002 (Tremblurel. 21, Last annotation update)				
DE KIAA0390 protein.				


```
N KIAA0390.
N Homo sapiens (Human).
C Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
C Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
N NCBI_TaxID=9606;
N [1]
P SEQUENCE FROM N.A.
P TISSUE=BRAIN;
C MEDLINE=97349984; PubMed=9205841;
X Nagase T., Ishikawa K., Nakajima D., Ohira M., Seki N., Miyajima N.,
A Tanaka A., Kotani H., Nomura N., Ohara O.;
A "Prediction of the coding sequences of unidentified human genes. VII.
T The complete sequences of 100 new cDNA clones from brain which can
T code for large proteins in vitro.";
L DNA Res. 4:141-150(1997).
L -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
R EMBL; AB002386; BAA20844.1; -.
R HSSP; P15822; 1BBO.
R InterPro; IPR000822; Znf_C2H2.
R Pfam; PF00096; zf-C2H2; 10.
R PRINTS; PR00048; ZINC_FINGER.
R PRODOM; PD000003; Znf_C2H2; 1.
R SMART; SM00355; Znf_C2H2; 10.
R PROSITE; PS00028; ZINC_FINGER_C2H2_1; 6.
R PROSITE; PS00157; ZINC_FINGER_C2H2_2; 8.
R DNA-binding; Metal-binding; Nuclear protein; Zinc-finger.
Q SEQUENCE 1300 AA; 14146 MW; 858B5B7047653664 CRC64;
```

```
Query Match 6.38; Score 330.5; DB 4; Length 1300;
Best Local Similarity 20.4%; Pred. No. 3.6e-11;
Matches 193; Conservative 94; Mismatches 303; Indels 355; Gaps 36;
```

```
DB 504 ERLQAAAKAEMDPVNSYOAMQLMARGMAHEGFLSKENHPLQRNHEDTLANAGVLPDKK 563
QY 772 EEDV-----TDEDLAGR 784
DB 564 REYLVGADGSKOKMPADLVHSTKVGSRDLPSKLDPLESSRDLFSLHGILQVLEYNLQCP 623
QY 785 GS-----ESGEKAISVRDSEFASGAEEVGY 813
DB 624 GNMKEKPTPCDGRVFRTHYGVVHSRVHKKRDKGEEDGLHVGLDERGSGSDQSQSV 683
QY 814 AAATAGKEMDENKTTQSSLPFRPPDDSLDQPPMNGSSGVLGKREGKPEPS--- 870
DB 684 SRSTPGSSNVTEE-----SGVGGLSGTGAQEDSPH 716
QY 871 -SSPASLTPE--GEATSVTLVEELSLQEMRKPEGSSSRKACECGQAFPSQALLEE 927
DB 717 PSSPSSSDIGEEAGKAGVQ-----QPALLRDSLSGAMKDCYCGKTFRTSHLKVH 769
QY 928 QKTHKEGRLFTVCFCRQGFLEKATLKKHMLAHQVQFPADHP 972
DB 770 LRHTGEKP-YKCPHCDYAGTQSALTKYHLERHHERONGA--GP 811

RESULT 27
Q9H8L4 PRELIMINARY; PRT; 498 AA.
AC Q9H8L4;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE CNFA FLJ13479 fis, clone PLACE1003738, weakly similar to zinc finger
DE protein 135.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=PLACENTA;
RA Isegai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Mishikawa T., Nagai K., Sugano S., Takahashi-Fujii A., Hara H.,
RA Tanase T., Nomura Y., Togiya S., Komai F., Hara R., Takeuchi K.,
RA Arita M., Nabekura T., Ishii S., Kawai Y., Saito K., Yamamoto J.,
RA Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y., Oshima A.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
C EMBL; AK023541; BAB14602.1; -.
DR HSSP; P25490; 1UBD.
DR InterPro; IPR000822; Znf_C2H2.
DR Pfam; PF00096; zf-C2H2; 13.
DR PRINTS; PR00048; ZINC_FINGER.
DR PRODOM; PD000003; Znf_C2H2; 2.
DR SMART; SM00355; Znf_C2H2; 13.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 13.
DR PROSITE; PS00157; ZINC_FINGER_C2H2_2; 13.
KW DNA-binding; Nuclear protein; Zinc-finger.
SQ SEQUENCE 498 AA; 54872 MW; C704FCF55CDAE1A CRC64;
```

```
Query Match 6.28; Score 329.5; DB 4; Length 498;
Best Local Similarity 22.0%; Pred. No. 1.3e-11;
Matches 146; Conservative 70; Mismatches 240; Indels 209; Gaps 24;
```

```
QY 352 SGELSYGEVWGLEKPGRRKRCRCAKVFGSDSAOIHLRSHTGPERYKCNVGNRFTTR 411
DB 14 AGELEP-----RCAHCPRAYGALSKLKHORGHTGEPKYACADCGKSTADP 59
QY 412 GNLKVHRRHREKYPHYQMNPHVPEHLDYVITSSGLPYGMSVPEKAEAEATPGGVE 471
DB 60 S-----VFRKHRR--TAGLHP-----YSCERC GKAYA-ELKDLRNHRSHY-----G 99
QY 472 RKPLVAST--TALSATESLTLTSTAGTATAPGLPAFNK-FVLMKAVEPKNKADENTPPG 528
```


RX MEDLINE=98087435; PubMed=9427563;
RA Miyake J.H., Szeto D.P., Stumph W.E.;
RT "Analysis of the structure and expression of the chicken gene encoding
RL a homolog of the human RREB-1 transcription factor";
CC Gene 202;177-186(1997).
DR EMBL; AF013754; AAB9584.1; -.
DR HSSP; P08045; 12NF.
DR InterPro; IPR000822; znf.C2H2.
DR Pfam; PF00096; zfc2h2; 15.
DR PRINTS; PR00048; ZINCFINGER.
DR SMART; SM00355; znf_C2H2; 15.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 13.
DR PROSITE; PS50157; ZINC_FINGER_C2H2_2; 13.
DR DNA-binding; Metal-binding; Nuclear protein; Zinc-finger.
SQ SEQUENCE 1615 AA; 178279 MW; 5C2F5F0C0FA16C79 CRC64;

Query Match 6.1%; Score 322; DB 13; Length 1615;
Best Local Similarity 18.9%; Pred. No. 1.5e-10;
Matches 268; Conservative 153; Mismatches 457; Indels 538; Gaps 57;

QY 4 ESEBSRLGVPAGEPRLG---GDASEED--HPOYCAK----- 36
DB 160 EMDREER--TPRAKVEDGQYGEGRKEDDAYHCVCFKFECKYALESHMETHPNLSLR 217
QY 37 ---CCAQFTDPTEFLAH---QNACTDPYVAVITGGQENPN-----NSSASSE 78
DB 218 CDICCTFTFTHRLRLHNAVIHKLPRDPTGKPLF--QNNPSIPAGFHDLGTFDFSCRF 275
QY 79 PRPEG-----HNNQVMDTEHSNPDGSSV 104
DB 276 PRISQVWCETNLRCISEFHRIETGCKNAKFAMLLALKLHTEHVMDQGRDKHKLSTSL 335
QY 105 PRDPTWGPERRRGESSHFLVAATG-----TAA 132
DB 336 PSE---NPQKA-----FMASLGLOYTKDKPKVAKOEDTODEVOEMRLALKSNLPQE 385
QY 133 GGGGGLILSPKGLAPLPPESTRAPPPPPPPGVSGHNLPILELHVLQOROI 192
DB 386 PSTGTLISLPLEAATMGPPFV-LPPTKE-----NIKLL-SLPDFQKGF 429
QY 193 HQMQMTEQI-----CROVLLGSLGQTVGAPASPELPGTASTTKPLPL 239
DB 430 IQPDDSSIVAKPISNESAIELADIQILKMAS-----SAPQIIS-LPPLSKAPSV-PVQSI 482
QY 240 FS---PI--KPVQTSKTLASSSSSSSSSSSGAETPKQAFHLKHPLSQHPFSAGVGRSH 294
DB 483 FKHMPRLKRPPLVTPRTIVAVTS-----TPP-----PLISAQOASPGCI--- 520
QY 295 KETPASPALPGSTDQIL-----ASPHLAFPSYTGLLAOCIGAAR-----G 336
DB 521 -----SPSLPPPLRLIKNSVEISSNSHLSOPGAKSSPSSQILLDPKVEPLQIHEKQTQ 574
QY 337 LEATASPGLLKPKNGSELGYEWMGLEK--PGGRHK-----CRGCAKYFGSD 383
DB 575 LEQDSTIEALLPLINMEAKIKQETEGDLKAITAGAAANKKAPTMRKVLYPCRFCDQVAFS 634
QY 384 SNAQIHLRSHTEGRRYKCNVC-----GNR-FITR-----GNLKYNHFH 419
DB 635 GYLRAIHRSLHGISPYOCNTCDYLAADKAALIRHLRTHSGERYICKICHYPTVKANCE 694
QY 420 RHREKPRHYQMNPRVPEHLDYVITSSGLPYGMSVPERKA-----EEBAAT 465
DB 695 RHLRK-KHLKLVTRDKIEKNIEYVSNAAEMDAFCSPDVCKLGEDLKHVRLAIRIMRT 753
QY 466 PGGVERKP-----LVASTAATASLES 489
DB 754 HSGCOKKKRPECKEGCTAFSAKRNCIHHILKOHLYQEREIENYIIVDSCAQSHTDAP 813
QY 490 LSTSGATATAPGLPAF---NKFVL-----MKA 514
DB 814 LLEDSTYMDCKPLTPLEDPQNGFSLGTSSHVPITKLEPMGNFPMDFDEPLDPSOKSKNL 873

QY 515 VE-----PRNKADENTPGSE----- 530
DB 874 VQVKQENLVSPLSFYDCSMEPIDLSIPKYLKRDNDIPGARNOELASSVITDNVNCQ 933
QY 531 ---GSAISVAESSSTAT-----LMQ 547
DB 934 QCPILGFGANGNSBKRAVGHDPRLKGSLLHVPITISPLLGNALLRPLRPKPPQPLP 993
QY 548 LSKLMTSLPFWALLTNFKSGSPFLPCARALGASPSFTSKLOQLVEKIDROGAVAVTS 607
DB 994 KPPVTKELPPLASTIAQIISVSSAPALIKTEAADAAPRAASSSTGCDKGNAKAKMTIVT 1053
QY 608 AA---SGAPT--TSA--PAVSSSASGPNOCVTLRVLSCPALRLHYONGGERPFCK 660
DB 1054 AIGRDSNLPEDLIQACDPEPSIADTG-----LTGRGRKKGTKN-KPK 1096
QY 661 VCGRA-----FSTRGNLRAHFVGHKASPARA-----QN----- 689
DB 1097 LSSGVLDSESGEFASIKKMLATTDINKFSPPLQSTDNFKESGQNGTSEDEKETPEDKL 1156
QY 690 -----SCPICOKKFTNAVTLQOHVRRHNLGGQIPNGGTALPPEGG----- 728
DB 1157 RGRNNTYSDCPQKTTCTCYCPRVFSSWASSLQRLMLTHDSQ---ADTEAPATGGEVLDLTS 1213
QY 729 AAQENGSEOSTVSGAGSPPOQSOQSPD--EBELSEEEDEDEEED----- 774
DB 1214 CEKEQPEVESELPESECSPOEQKADSPAEDEAEKAEDEYEGREDPSVNSKSLDNFA 1273
QY 775 -----VTDEDSLAGSGESGGEKATSVRGDSEERSCA--EEEVGTVAATAATGKEMDS 825
DB 1274 SKLMEFKLASDQSGASSQTERKHACDVCQKTFKFGALSRKKAHIREDRKDERSED 1333
QY 826 NEKTOQSSLPPEPPPSLDQPO--PME--QSGSGVLCGKEEGKPERSSPASALTPEGEA 883
DB 1334 ESKSIDDAQAPSMQDSGLEBESPMDLKVYSPPLDCEATGKNESES-----ISGEG 1388
QY 884 TSVLVEBELSLQEMRKREPGESS-----RKACEVCGAAPPQAALEHQTKHPKEGPLE 938
DB 1389 TERKSTESKSDDKIPKIDKASTAKADKRRKVCYCNKRKFWSLQDILRHMRSHTEGEP-Y 1447
QY 939 TCYFCRQGFLEKATLKKHMLAHNOVOPRAPGPN 974
DB 1448 KCQTCERTFTLKSLVRRQRI-HQKVKNTRNHGKES 1482

RESULT 31
Q96I71
ID 096I71 PRELIMINARY; PRT; 734 AA.
AC 096I71;
DT 01-DEC-2001 (TReMBLrel. 19, Created)
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Unknown (protein for MCC:12654).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=UTERUS;
RA Strausberg R.;
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC007777; AAH07777.1; -.
DR InterPro; IPR003309; Treg_SCAN.
DR InterPro; IPR000822; znf_C2H2.
DR Pfam; PF02023; SCAN; 1.
DR Pfam; PF00096; zfc2h2; 13.
DR ProDom; PD000003; znf_C2H2; 6.
DR PROSITE; PS50804; SCAN_BOX; 1.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; UNKNOWN_13.
DR PROSITE; PS50157; ZINC_FINGER_C2H2_2; 13.
KW DNA-binding; Zinc-finger.
SQ SEQUENCE 734 AA; 82055 MW; B804766A0B08048 CRC64;

	Query Match	6.18:	Score 321;	DB 4;	Length 734;	
	Best Local Similarity	20.7%:	Pred. No.	6.6e-11;		
	Matches 196;	Conservative	74;	Mismatches 309;	Indels 368;	Gaps 37/;
Y	80 REGNHNPVMDT-----EHSNPDSSGVPTDTPGPERREESSGHNLVAATGAAG	133				
b	74 REVNSKEQMLLELVLEQFLGALPPEIARYO-----GQRGSPBEAALVDGLRRERPG	127				
Y	134 GG-----GGLILASPKLIGAT--PLPESTAPRPPPPPBPVGSGHLNIPILE	182				
b	128 GPRRWTVVOVGGEVLSEKMEPSFQPLPETERPPEPKPTP-----	171				
Y	183 ELRVLOQRDI-----HQMTQEIQICROVLLLSLQGYVAPASPSLPETPGCTASTPRL	237				
b	172 --RTMQESPGLGVKKESSEVED-----SDPLESGPIAAQTGSV	208				
Y	238 PLFSPIKPQVTSKTLASSSSSSSSGAEFPKQAFENLHPGLSOPFGSAGVGRSHKPT	297				
b	209 PLLPBEARCGIVL--DLFRPHSKTGEGBSWR-----EHPRALMHE--ENGITF--	255				
Y	238 PAPSPAL-----PGSTDQLASPILAPSTTGL--LAADCLGARGLGEATASPGLLK	347				
b	256 -SPGVALQGISISAGESS--YSPHLHVWDLMAGLSGIQSISR--EGGFANLLL	307				
Y	348 PKNGSELVYEV-----MGPL-----EKGGRH-----KCRCAKAFGS	382				
b	308 PSDLSSEDPDTDDPCRCGVGPALLTTTRMSRPGRGRGRPSTGGVYRGGCIDVCGRVSQ	367				
Y	383 DSAOLIHLSHTGERPYCKVAVCGNRFTTGNLKVFNHRREKYPHYOMNHPYENDIYY	442				
b	368 RSNLLRHOKIHTEGRPFVSCGCSGFSRSSHLLRMQLHTTERP-----FV	413				
Y	443 ITSSGLEPGMSVPERKAEEEAATGCGVERPVLVASTALSTATESIITLISACTANAPG	502				
b	414 CGDCGGGF--VASARLEEHRRVHTG---EDPFCACGQSQRORSNLLQHORIHGDPG	467				
Y	503 LPAPNFVIMKAVEPKKNKADENTPPGSEGSISGAVESTATLMQSLIKLTSLPSMALLT	562				
b	468 -----PGAKPPRAPGAP-----	480				
Y	563 NHFKSTGSPPLPLCARALGASPSSETSKLDQVEKIDROGAVAVTSAASGAPTSPAPSS	622				
b	481 ---PPGPPPCSCTR-----ESPARRAVLLEHQAVH-----TG	509				
Y	623 SASSGPNOCVICLRVLSCPALRLHYOHGGERPPKCYKGASAFSTRGNIRAFVGHKAS	682				
b	510 DKSFQ---CVEGGERGRRSVLLQHRVRHSEGRPAACAECGGSFRRSRNTLOHRRITGE	566				
Y	683 -PARAONSCPICQKFTNAVTLIOHHVMHLGGQIPNGTALPEGGAQAENGSEOSTVS	741				
b	567 RPFA-----CAEGKAFRQRPTYLQHLRVNHGEK--PFACPCEG-----	603				
Y	742 GAGSPPQQSQDSPEBELSEEEEEDEEBEDVDDEDLSLAGRSSESGBKAIIVRGDSE	801				
b	604 -----QRFSSQIKLTRHORHT-----GBK-----	623				
Y	802 EASGAEEFYGAANAATAGKEMDSNEKTTQSSLPPRRPDSLDOQPDMQSSGVLGK	861				
b	624 -----RYNHGEGGL-----	632				
Y	862 EEGKPERSSPASALTPEGEATSVTLVEELSLDEAMRKPESSSRKACEVGOAPPSQ	921				
b	633 -----GFTQYSRLT--EHQRHTHERPF--ACPECGQSFRQH	665				
Y	922 AALEHQKTHPEKGPLTCVFCROGFLERATLLKHMILLANHVOQRA	968				
b	666 ANLTQHRRIHTGERP-YACPECGKAFQRORPTLQH-LRTHREKPRPA	710				
RESULT 32						
D 924S6	PRELIMINARY:			PRT: 726 AA.		

Query Match	Best Local Similarity	Matches	Score	DB	Length
Query Match	Best Local Similarity	Matches	Score	DB	Length
1	1	1	1	1	1
2	2	2	2	2	2
3	3	3	3	3	3
4	4	4	4	4	4
5	5	5	5	5	5
6	6	6	6	6	6
7	7	7	7	7	7
8	8	8	8	8	8
9	9	9	9	9	9
10	10	10	10	10	10
11	11	11	11	11	11
12	12	12	12	12	12
13	13	13	13	13	13
14	14	14	14	14	14
15	15	15	15	15	15
16	16	16	16	16	16
17	17	17	17	17	17
18	18	18	18	18	18
19	19	19	19	19	19
20	20	20	20	20	20
21	21	21	21	21	21
22	22	22	22	22	22
23	23	23	23	23	23
24	24	24	24	24	24
25	25	25	25	25	25
26	26	26	26	26	26
27	27	27	27	27	27
28	28	28	28	28	28
29	29	29	29	29	29
30	30	30	30	30	30
31	31	31	31	31	31
32	32	32	32	32	32
33	33	33	33	33	33
34	34	34	34	34	34
35	35	35	35	35	35
36	36	36	36	36	36
37	37	37	37	37	37
38	38	38	38	38	38
39	39	39	39	39	39
40	40	40	40	40	40
41	41	41	41	41	41
42	42	42	42	42	42
43	43	43	43	43	43
44	44	44	44	44	44
45	45	45	45	45	45
46	46	46	46	46	46
47	47	47	47	47	47
48	48	48	48	48	48
49	49	49	49	49	49
50	50	50	50	50	50
51	51	51	51	51	51
52	52	52	52	52	52
53	53	53	53	53	53
54	54	54	54	54	54
55	55	55	55	55	55
56	56	56	56	56	56
57	57	57	57	57	57
58	58	58	58	58	58
59	59	59	59	59	59
60	60	60	60	60	60
61	61	61	61	61	61
62	62	62	62	62	62
63	63	63	63	63	63
64	64	64	64	64	64
65	65	65	65	65	65
66	66	66	66	66	66
67	67	67	67	67	67
68	68	68	68	68	68
69	69	69	69	69	69
70	70	70	70	70	70
71	71	71	71	71	71
72	72	72	72	72	72
73	73	73	73	73	73
74					


```

Db      356 AVEPLPALLIAPAPAPAEERREPPSLIGLISVRAGEVVRPNAGEADPGGGRSYGGERLPDS 415
Jy      795 -----SVRGSEEAASGAEEEVYTAATAAGKEMDSNEKTQQSSLPPEPPDSDLOPOP 849
          : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      416 ALPNRRARRHREEDEEBEVEVAEEESMARGRSLGS-----LTSLNHP----- 459
Jy      850 MEDQSSGVLGKEEGKGKPRRSSPASALTPREGATSVTLVELSLQEAAMKEPEGESSRK 909
          : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      460 -----GEGSGCPAPAAGTQANSTA--TOENGLLVGGTRSRAGRATGK 501
Jy      910 ACEVCGAFAEQALLEEOKTHPKKEGRLEFCVCROGFLEAKTKHHMLAHNOV---- 965
          : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      502 DCPFGCSFSAHHKLHLRVHTGERP-YKCPHDVACTOGSKLYH-LQRHREORSSA 559
Jy      966 ----PFAPHGPONIALSLVPCCSPSITS 990
          : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      560 GPGPPPEPPPSQSGSLDPQSAGKKPTQAS 588
          : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

RESULT 33
P91365 PRELIMINARY; PRT; 2232 AA.
P91365
AAC P91365:
CUT 01-MAY-1997 (TREMBLrel. 03, Created)
CUT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
CUT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
DET K0649.1 protein.
JN K0649.1.
CS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabdilitida; Rhabditoidea;
OC Rhabdilitidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL NZ;
RA Geisel C., Gattung S.;
RL Submitted (JUN-1987) to the EMBL/GenBank/DDBJ databases.
-1- ALTERNATIVE SPLICING OF THE SAME GENE. THE SEQUENCE SHOWN IS THAT
OF THE A FORM.
CC EMBL; U08046; AAC70889.1; -
CC EMBL; U08046; AAC70890.1; -
KW Alternative splicing.
FT VARSPIC 842 866
FT FT
FT VARSPIC 870 1051
FT FT
FT VARSPIC 870 1051
FT FT
FT FT
FT FT
FT FT
FT FT
FT FT
SO SEQUENCE 2232 AA; 213840 MW; 08D69FR638E14CC8 CRC64;

Query Match 6.0%; Score 316.5; DB 5; Length 2232;
Best Local Similarity 21.2%; Prid. No. 4.5e-10;
Matches 231; Conservative 128; Mismatches 435; Indels 295; Gaps 48;

OY 51 ONACSTDPPVMVILIGGOENPNNS-----ASSEPRPEGHNNPVQVDTEHSNPDDGSSV 104
          : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 498 QASAFSPETMSTVSG---PTGSTIVYVPGSGTSPAPSSPNPS-----SSFASGTSTI 548
          : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
OY 105 PTDPWTGPERGEES-----SCHFLVAATGTAAAGGGGLIASPKIGAMP-----LPPE 153
          : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 549 TIS-----GSSITIVSVSGSTVGSGTGNQS-----TLASTT-AVPGSSVYVPS 593
          : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
OY 154 STPAAPPPIPPP-----PPPGVGSGLINPLILEELKVLQGRQIHQMQ--- 196
          : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
594 SSPQSSGSPAPVTGTTSPQSQTSSQSPSPSNPPS-SYPTGSSQSTITTPGSTRASSPTGS 652

```

QY	197	-----MTEQICROVLLLG-SLG-QYVGAPAPSEEL-----PGTGTAAS	232
Db	653	TGSTFSAVTEVTSQGTVPDGSGLGQNSTMSPSLSLSPSTSGMSTLTSEPSSTGSS	712
QY	223	TKPLPLPSPIKPVOTSKTLASSSSSSSSSGAEPLPKQAFPHILHPLGSHPPSAGVCR	292
Db	713	AQSTLTTPSP-NPQSSTSLSESTSGATSSSGAGTTMT-----SPQSSSVAS	760
QY	293	SHKPPADSPALPGSTDOILASPHILAFPTGTGLLAOCGARGLEATAPG--LLKPKN	350
Db	761	SQGST----SPASSTTSGEMTSQGSQTQPCSSVSTSAILLTSTQGSVSTMSPGSVTRPST	817
QY	351	GSGELSYGE--VMGLEPKGRHKRCFKACVFGSDSAILOILKSHTGEPHYKCNVCGNRF	408
Db	818	VSGSTSSGSTVVGSTPE-----ASTGSSVA-----	843
QY	409	TTKGNLKVHFRHREKYPIVQNMHPVPRPHLDVYTSSGLPFGMSVPREKAEDEATPGG	468
Db	844	-----SSSPAPSTSONPNPNTSSGSSMTQSPYPSOSTSP---VESSITPPSP	887
QY	469	GVERPPIVASTALASTEILLTSLTASGATATAPGAPFKFVLMKAVEPKNADEMTPPG	528
Db	888	GSPGTTL--TSTSPBSQSTTTGISTQG--STSPGISTTS-----EEMTSQGSQTRG	935
QY	529	SEGSAT---SGVAES-----STATLQSLKMTSLPMSALLTNHFKSTGSPPLPCARALG	581
Db	936	STGSGTYQPSVYSDSTSSSGSTVYGTSEBSSSPILPSTQONTNPTSSGS-----	984
QY	582	ASPSSTKIQOLEKIDROGAVAVTSAAGAP-----TTSAPAPSSASASGPNOCVLCRY	637
Db	985	SMSTQPOQSOSTSPVESSTSGATSS--SGSPGTTLTSLSPSPSPSTIGSSQGSTSPVY	1042
QY	638	LSCPRALRLHYOHGGERPFKCKVCGRATS---TRGNLRAHFV-----GHKASP	683
Db	1043	STISQGSTPTPGSTGTYTKRPTSVSSASASGATATMGSTBASSTSGGSSSTPRPSQSTSP	1102
QY	684	AARAOISCPICOKKFTNAVTLQOHVBMHLG-----QIP--NG	719
Db	1103	STSGATSSPGSGSTTLTSLSPSQSGSTIGSGGSTSPVYVTTSGDMTSQGTQIPGSTG	1162
QY	720	GTALPRGGAGADEGSEOSTVSAGSFPOQSOQPRPEELSEEBEEDDEEDYDDED	779
Db	1163	STVYTOPSTSGSTSTGELTISQSTQTP--KSLSTSPALSTSTQGSVSTNPSGTYTPQS	1221
QY	780	SIAGRGSESGEKAIVSRGDSEFASGAEEVEYVAAATGKIMDSNEKTYQOOSLPPP	839
Db	1222	TV--KSTSSG-----SVYTGSTGSGSTSGSSMATSLSS--SSP	1257
QY	840	PPDSLDQOPMEGSSGVLGKEEGGKPERSSSPASALPRPEGATSVTLVELSLQEAHR	899
Db	1258	VPTSGSPNPSTSGSS-----PPTPNPQOSTSPVVS--TTTGEMTSHGSTQTPSTIGSTV	1310
QY	900	KEP-----GESSSKRAEY-----CGQAF---PSQAALBEHQKTHKBEPLFCVBCRQ	945
Db	1311	TOPSTVSGSNSSGSTVTTGSSSEASTSGSSKRTSPSSIS-----EYPTSSPISPTTFASS	1364
QY	946	GFLERATLKKHMLLAHNOVQFPAHPGQNTAALISLVPGCSPSTST-----GLSPFR	998
Db	1365	--TSGSTISDVSSVSTTSLARLSSLP-----STVPSSTQPSSTSEGGSKKASSPVPS	1416
QY	999	K--DDPTIP	1005
Db	1417	QTSSTPTNP	1425
RESULT 34			
ID	061360	PRELIMINARY;	PRT; 962 AA.
AC	061360;		
DT	01-AUG-1998 (TREMblrel. 07, Created)		
DT	01-AUG-1998 (TREMblrel. 07, last sequence update)		
DT	01-JUN-2002 (TREMblrel. 21, last annotation update)		
DE	CROL alpha;		

CROL OR CROL ALPHA OR CG14938.
Drosophila melanogaster (Fruit fly).
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephyroidae; Drosophilidae; Drosophila.
NCBI_TaxID=7227;
(1)
SEQUENCE FROM N.A.
STRAIN=CANTON S;
D'Avino P.P., Thummel C.S.;
Submitted (Aug-1997) to the EMBL/GenBank/DBJ databases.
-1- SUBCELLULAR LOCATION: NUCLEAR (By SIMILARITY).
EMBL: AF020347; AAC1516.1; -.
HSSP: P07248; 1ARD.
FlyBase: FBgn0020309; crol.
InterPro: IPR000822; znf_C2H2.
Pfam: PF00096; zf_C2H2; 18.
PRINTS: PR00048; ZINC_FINGER.
ProDom: PD000003; znf_C2H2; 5.
SMART: SM00355; znf_C2H2; 18.
PROSITE: PS00028; ZINC_FINGER_C2H2_1; 18.
PROSITE: PS50157; ZINC_FINGER_C2H2_2; 18.
DNA-binding: Metal-binding; Nuclear protein; Zinc-finger.
SEQUENCE 962 AA; 106360 MW; 618D1BC9799C9AAB CRC64;

Query Match 6.0%; Score 314.5; DB 5; Length 962;
Best Local Similarity 19.3%; Pred. NO. 2.2e-10;
Matches 202; Conservative 94; Mismatches 405; Indels 345; Gaps 36;

48 LAHONAGSTDPVPM-----VILGQENPNSSASSERPREGHNHPQVMDTEHSNP 97
1 MQHVSASSSVSVTPVVTGTTGTTGPPPLPKSEHKEGKP-----PHGEMKYKNI 55
98 PDGSSSVPTDPTWGPERRGESESGHFLVAATGTAAGGGGHLASPKLGATPLPRESTPA 157
56 EDISQLFLYHEVFG-----KHGDVNVNHQLAAHGGQL----- 88
158 PPP 217
89 ---PPPPPLPPQVTS-HAASAAAAAATASTNMAVAAVMAAANAAAAAASAGG--GL 142
218 PPSPELPGTGAASSTKPLPLPSPKIPVOTSKTGLASSSSSSSSSGATPPQAFFHLVH 277
143 PPAIS---GNGG-----QQVTVTSSSTSSGSGTSSGTTTAGELLMPK 185
278 PLGSQHPSA---GGVGRSHKPTAP--SPALDGS-----TDOLASPLHAF 320
186 MEGGIHGVDSGNGGQVVALADGPTATGTHVCDIGKMFQRYQLIV--HRRVH 243
321 STTGLLAQCIGAANGLETASBGLLKPKKSGELSTGEVNGP-----L 364
244 SERKPFMCQVCGQG---FTTSQDLTR---HGKIHIG---GPMFTCIYCFNVFANNTSL 292
365 EKPGGRHK-----CRFCAKVGSDSALQIHLRSHGERPRYKCANVCGNFFTRGNLKVH 418
293 ERHKKRSHDPRFACTICOKTARKKHLNDNHFSSHGTETPRCOYCATFTTKREHVNHV 352
419 HRHREKYPH---VQNNPVRPETHL--DYVITSSGLPYGMSVPRPEAKEEBAATPGGVERK 473
353 RKHTGGERPHRDICKKSPFRKEHYVNHVMHTGGTTPHOCVCGKK-----YTRK 401
474 PLVASTALSTESTLSTLSAGTATAPGLPARPKFVLMKAVEPKNKADEMTPRGSEGA 533
402 EHLNHRKSHNET-----PRCEICGKSGFSRKEH----- 431
534 ISGVAESSTATLMQSKLMTSLPSWALLTNHFK-STGSFP--LPLCARALGASPSSTSKL 590
432 -----FTNHILMHTGETPRHRCDCSKTF-----TRK 457
591 QQLVEKIDROGAVAVTSAAGAPTTAPAPSSSSASGPNOCVCLRVLSCPRALRLHYGO 650
458 EHLNLNH-RQ-----HTGESPRHRCYSCKMTFTTKREHVLVNHINO 494

QY 651 HGBERPEKCKVQGAFTSRGNLRHFVGHKASPAARAONSCPIQCKPPTNAVLQOHVM 710
495 HTGTPRFCKTYCTKFAFTRD---HMVNVHQRQHTGESPRHKCTYCTKFTTKREHVLNHVQ 550
QY 711 HLGGQITPNGTALPRGGGAQENGSEDSITYSGAGSFQOQSOQPSPEBELSEEEDEDE 770
Db 551 H-----TGDSPHRCYC-----KRTPT 567
QY 771 EEDPTDEDLSLGRSGESGGEKAISVGRDSEASGAEDEVGTVAANAATAGKEMDSNEKT 830
Db 568 RKEHLTNVRLH-----TGDSPHRCYCQKTT-----RKHNLNH--M 604
QY 831 QQSSLP PPPPSDIDOPRMEQSSGVLGCKEKGKDERSSPASALPREGATVTLVE 890
Db 605 RQHSNDPNHCNCVKRPPTRKHLNLNHSRCHGDRP-----FTCEGCKSPFLKG 655
QY 891 ELISQEAARKREPGSSSKRACEVCGAFTSQAALEENOKTHPKRGRLFTVCRCGFLE 950
Db 656 NLLEFHQNSHTKGQEMERPFACBCKPKNFLCKGHLVSHMRSHSGKRP-HACTLCSKAVER 714
QY 951 ATKKHMLAH-----HOVQFA 968
Db 715 GNLRKHKMKNPRDAMPPRPVPHRQIPAGVLTQVQGVKPIIIPHSATTTMHTIQOLT 774
QY 969 PHGPNTAALSLVPGCSPTSTGLS 994
Db 775 AGAAGAGAVOLTPLGLVPLVTSTLLS 800

RESULT 35
070162 PRELIMINARY; PRT; 814 AA.
ID 070162;
AC 01-AUG-1998 (Tremblrel. 07, Created)
DT 01-AUG-1998 (Tremblrel. 07, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Myeloid zinc finger protein-2.
GN Zfp98 OR MZF-2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98074810; PubMed=9413999;
RA Murai K., Murakami H., Nagata S.;
RT "A novel form of the myeloid-specific zinc finger protein (MZF-2).";
CC 1- SUBCELLULAR LOCATION: NUCLEAR (By SIMILARITY).
DR EMBL: AB007407; BAA25190.1; -.
DR HSSP: P08047; 1SP2.
DR MGD: MGI:107457; Zfp98.
DR InterPro: IPR003309; Treg_SCAN.
DR InterPro: IPR000822; znf_C2H2.
DR Pfam: PF02023; SCAN; 1.
DR Pfam: PF00096; zf_C2H2; 13.
DR PRINTS: PR00048; ZINC_FINGER.
DR ProDom: PD000003; znf_C2H2; 4.
DR SMART: SM00431; LER; 1.
DR SMART: SM00355; znf_C2H2; 13.
DR PROSITE: PSS0804; SCAN_BOX; 1.
DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; 13.
DR PROSITE: PS50157; ZINC_FINGER_C2H2_2; 13.
DR DNA-binding: Metal-binding; Nuclear protein; Zinc-finger.
SQ SEQUENCE 814 AA; 91667 MW; ED3CC6C02CD83871 CRC64;

Query Match 6.0%; Score 314; DB 11; Length 814;
Best Local Similarity 21.1%; Pred. NO. 1.9e-10;
Matches 200; Conservative 73; Mismatches 303; Indels 374; Gaps 39;
QY 80 RPECHNNPQVMDT-----EHSNPPDSSGVPTDPTW---GPER-- 114
Db 154 RPEVHSKEQVLELLVLEQFLGALPPEICAOVQOGQPGSPEDAAALVRLRLWELDGRPKMV 213

QY 115 ----KGESSGHFLVATGTGAAGGGGLIAPKLGATPLPESTPAP-----PPPPP 164
Db 214 TVQVQGEK-----VLSEKME-----PSSFQPRPHIKIQTDPDPAE 248
QY 165 PPPPGVSGHLNPILEELRVLOOQROIHOMQTEQICROYLLLSL--GOTVGAASPS 222
Db 249 TPQGMQGEPLRLPM-KEETELLGNPEL-----LESGLPAAQEAATLLPK 294
QY 223 ELPGTGTA-----SSTKPLPLFS---PIKPYQTSKTLAASSSSSSSSSGAETPKQA 271
Db 295 ETQGGHTLDTQTSPHSDTEBDVPWSEDPMLEHVEGVGIFSPGFTLQMDSTAE--PDM 353
QY 272 FPHLKHPR-----LGS-----QHPSAGVGGRSHKTPRAPSALP-----GSTDLIASPHLA 318
Db 354 SPHLHVPMWLDMAISLGOIQSPTEGPF-----AAHALVLPSPDGGEDPACEDP--- 402
QY 319 PPSSTGLLAACIGAAAGLEATATSPGLLKPKNGSGELSYGEVWPLEKPGGRHKRCFCAR 378
Db 403 CFSVGRALVATRNHAAHARGRSQS-----HPSGTCAMR-----GGR--CDVCGK 443
QY 379 VFGSSDALQIHLSHGERPYKCNVCGNREFTTRGNLKVHFRHREKYPHVQMNPHVPEH 438
Db 444 VFSQSRNLIRHQKIHTGERPFVCGECGRFSRSSHILRHQLTHTTEERP----- 491
QY 439 LDYVITSSGLPYGMSVPRPEKAEAEATPGGVERKPLVASTALSTESLTLLSTAGTA 498
Db 492 --FVCRDCCQGF--VRSALYEEHRRVHTG--EOPFRCTEGCGSFRKSNLLQHORIG 543
QY 499 TAPGLPAFKKFLVLMKAVERPKNAKADENTPPGSEGAISGVAESSTATLMOLSKLMTSLPSW 558
Db 544 DRPG-PA--PSILPLAGVPE-----PRGP-----FQSCCECESPRR 577
QY 559 ALLTNHFKSTGSPFLPLCARALGASPSETSKLQOLYEKIDROGANAAVTASAGAPTTSAP 618
Db 578 AVLLEH-----QAVHTGDK----- 591
QY 619 AASSASSSGNOCVICLRVLSRPARLRHYGOHGERPEPKVCYCGRAFSPRGNLRAHFG 678
Db 592 -----SFG---CVEGGERGRSVLLQHRVHSEGRPRACBEGGSFROKSNLITQHORI 642
QY 679 HKASPARAARONSCPTCOCKFTNAVTLDQNHVMHLGQIIPNGGATLPEGGGAOENGSEOS 738
Db 643 HNGERAF-----ACAEGCKTFRQATPLTQHLCHVTEK-----PRACBEGC----- 683
QY 739 TVSGAGSFPQOQSQRPSEELSEEEDEEEDVDYDEDSLAGSGESGGEKAISVRG 798
Db 684 -----QRFQORLKLTRHQRTH-----GEKPYCC-- 707
QY 799 DSEASGAEEVEGVAAATATGKEMDSNEKTTQOSSLPRPPRPSLIDQPMEGSSGVL 858
Db 708 ----- 707
QY 859 GKGEGKPKPSSSPASALPREGATSVTLVELSLQEAAMRKPESSSRKAGEVCGCAF 918
Db 708 -----GECDLGFTQVSRLT-----EHQRIHTGERPF--ACBEGCGSF 742
QY 919 PSCAALEENOKTHPEKGRPLTVCFCROGFLERATLKKHMLAHNOQPPA 968
Db 743 ROHANLTQHRRIHTEGERP-YACAECGKAFROPRTLQH-LRTHRHEKPPA 790
RESULT 36
ID 046205 PRELIMINARY: PRT: 1920 AA.
DT 01-JUN-1998 (TREMBLrel. 06, Created)
DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE Zinc-finger nuclear protein HINDSIGHF.
GN PEB OR HINDSIGHT OR EG:66A1.1 OR CG1212.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

OC pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydriodes; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97330681; PubMed=91871140;
RA Yip M.L.R., Iamka M.L., Lipsitz H.D.;
RT "Control of germ-band retraction in Drosophila by the zinc-finger
protein HINDSIGHT".
RL Development 124:2129-2141(1997).
CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
DR EMBL: U86010; AAB95640.1; -
DR HSSP: P07248; 2ADR.
DR FlyBase: FBgn0003053; peb.
DR InterPro: IPR000822; Znf_C2H2.
DR Pfam: PF00096; Zf-C2H2; 12.
DR PRINTS: PR00048; ZINCFINGER.
DR SMART: SM00355; Znf_C2H2; 13.
DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; 11.
DR PROSITE: PS0157; ZINC_FINGER_C2H2_2; 10.
KW DNA-binding; Metal-binding; Nuclear protein; Zinc-finger.
SQ SEQUENCE 1920 AA; 208973 MW; BA51E67584916AA2 CRC64;
Query Match 5.9%; Score 313; DB 5; Length 1920;
Best local Similarity 20.0%; Pred. No. 6e-10;
Matches 249; Conservative 137; Mismatches 438; Indels 424; Gaps 56;
QY 16 GEPAEIGDASEEDHPQVACACCAOFTDPTPEFLAHONACSTDPPVAVITIGQEPNNSS- 74
Db 382 GNLTLSAGAESYEED-----ASCs---TDVSSGSHSRSSS-----LNNNNNSH 423
QY 75 -----ASSPPRPGHNNPOVWDTEHSNPDGSSVPTDPTWGPERRGESSG 121
Db 424 KANNNLKDLBELEVSTEDQDTEENKORILKTTINNNTIEEQQEDMD-----DEADADV 478
QY 122 HFLVAATGTAGAGGGGLIAPSKLGATPLPRESTRAPPPPPPP-----PGVSG-- 173
Db 479 AMLTSTPDVAT-----LLAGASASGA-----ASRPPTSPSPASALLTSCPACGASDFE 527
QY 174 -----HLNPILEELRVL-----QORQHOMQTEICR-QVLLLSLGQTVG 216
Db 528 TLRALCVNLAMHSDIPRKCRCCEVIFATNRQDSH-----CRLRYMLAGGLPRLG 580
QY 217 APASP--SELPGTASTKPLPLPSPIKPYQTSKTLAASSSSSSSSSGAETPKQAFH 274
Db 581 ASSSPPLNEPDEDEHGDDEL-----EQKRLASQSD----- 614
QY 275 LYNPLGSQHPFSAGVGGRSHKPTP-----APSPALPGSTDOLIA 313
Db 615 FPHQLYIKHKTKANCGAISHPPSPIKHEPADTKDLADIGILNMTSSSSSFLRNFQSVN 674
QY 314 SPHLAFPTTGILLAAOCIGAAAGLEATASPGL-----LKP 349
Db 675 TPNSQVSLDG-----RDQEEADDAFTSEFRMKLRGEFPKLCITAVPNNRLAL 724
QY 350 NGSGELSYGEVWPLEKPGGRHKRCFCAKVFGSDSALQILHRSHTGGRPKCNVCGNREFT 409
Db 725 KGNHRVILGAV-----GAPGPRRCNMCPYAVCDKALVLRIMRTHNGRPRPCACNVAFT 779
QY 410 TRGNLKVHF-HRH---REK-----YPHVQMNPH 434
Db 780 TKANCERHLRRNHGKTSREEVKRAIYVHPAEDAGCEDSKSLGEDLADTSFRSISPPRP 839
QY 435 VP-----EHLDVITSSGLPYGMSVPRPEKAE-----EEAATPGGVERKPL 475
Db 840 PPNVESKSQLKHM--LLENHNLAPVNOQPPKLTQVKSLLDQLVDKPPAPADQOQOQOQO 897
QY 476 VASTTALSATESLTLTSTAGTATAPGLPAFNKFLVLMKAVERPKNAKADENTPPGSEGAIS 535
Db 898 EKSGSALDFSNDVLDLSKKP-TGASLSLTPAVYTRTPPAVAVP-----TPGGVGTPLDA 950
QY 536 GVASSSTATLQO-----LSKIMTSLPSWALLTNHFKS-TGSPF-LPLCARALGAS 583

Db 951 AAIEQQOLLIAQQOLFAGGEGYMOOLFRLS-----MFOQSOTSGFPPEPFMA---PPP 999
Y 584 PSESTKIQQOLVEKIDRQGAVAVTAASGAPTTASAPSSASSGPNOCVLCRLVSCPR 643
Db 1000 PQANPKPMPVSPFNRLNPMV--GVGVVPP-----PGGFPKMYIKNGVL-HPKQ 1048
Y 644 LRLHYQHGGGERPPKCVCGRAESTGRLRAHF-----V 677
Db 1049 KQRRYPT---EREFPACHCSARFTLRSNMERRHVKKQHPQFYAQREKSAHHVMGRGASNV 1105
Y 678 GHKSPAARA-----QNSCRICCKKFTNAVLQOHVRM 710
Db 1106 AAAAAAATAATPVNAGGPGSSGFGSHHHGHGSHGHAPlTSEQ--VKCAILLAQOLKA 1163
Y 711 H-----LGGQIPNGTAL-----PEGGAQAENGSEOSTVSGAGSFQQOQSOOPS 755
Db 1164 HKNPTDLLQALANGSSSVACNPILHFGYPLPLNPSMHNHSSQ-----GNGQATMADDEPK 1219
Y 756 -----PEELSEEEDEEDEDVTDPSL-----AGKSESSEGE--KA 793
Db 1220 LIIIDEDENEHDHEVAEDVDDEFEDEDEEEMDEPEDELIDEQAEKEAEHEQELPK 1279
Y 794 ISVRGDSEASGAEEEGVTAATAAGKEMDSNEKTTQOSSLPPEPPPSLDOP---QPM 850
Db 1280 LEOLGTKEAAQKMAETI--LEQATKAGKPL--SPPTKKNASPAKPTVATTMOPAITAPS 1336
Y 851 EOGSS--GYLGGKEEGGK--PERSSSP---ASALTPEGE---ATSVT-----887
Db 1337 TNPSSLKTMTIAQAEVYKSLKEVASSPFPKDSQDLVPAKLVDMNTSQMMGFNSYRPSD 1396
Y 888 -----LVEELSLQF-----AMRKPEGSS-----SRKACEVCGQA 917
Db 1397 VANHMQSDEEGIVASGSASESNNSTGTEVDYTSSSSSSEPKKSAVSLADNRVSCPYCQRM 1456
Y 918 FFSQALAEHQKTHPKGRLFTCVFCRGGFLERATLKKHMLLAHQYO 965
Db 1457 FPMSSSLRRHILTHGOKP--FKCSHCPLLTFTTKSCDRHLKKHGNVE 1503
RESULT 37
D 19VKF1
C 09VKF1 PRELIMINARY; PRT; 934 AA.
N 09VKF1
Y 01-MAY-2000 (TREMBLrel. 13, Created)
Y 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
Y 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
N CROL protein.
N CROL OR CG14938.
Y Drosophila melanogaster (Fruit fly).
Y Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Y Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Y Ephydroidea; Drosophilidae; Drosophila.
N NCBI_taxid=7227;
N [1]
N SEQUENCE FROM N.A.
N STRAIN-BERKELEY,
C MEDLINE-20196006; PubMed-10731132;
X Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
A Amanalides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
A George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
A Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
A Brandon R.C., Rogers Y.H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
A Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Mklos G.L.G.,
A Adell J.F., Agbayani A., An H.-J., Andrews-Pfankoch C., Baldwin D.,
A Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
A Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
A Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
A Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
A Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
A de Pablos B., Delcher S., Deng Z., Mays A.D., Dew I., Dietz S.M.,
A Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
A Durbin K.J., Evangelista C.C., Ferraz C., Ferriere S., Fleischmann W.,
A Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,

RA Glodex A., Gong F., Correll J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Mei M.-H., Ibegam C.,
RA Jalali M., Kalush F., Karen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Piltman G.S., Fan S., Pollard J., Puti V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskaas R., Turner C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
DR EMBL; AE003633; AAF53121.1; -.
DR HSSP; P07248; 1ARD.
DR FlyBase; Fbgn0020309; crol.
DR InterPro; IPR000822; ZnF_C2H2.
DR Pfam; PF00096; zf-C2H2_17.
DR PRINTS; PR00048; ZINC_FINGER.
DR PRODOM; PD000003; ZnF_C2H2; 4.
DR SMART; SM00355; ZnF_C2H2; 17.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 17.
DR PROSITE; PS50157; ZINC_FINGER_C2H2_2; 17.
KW DNA-binding; Metal-binding; Nuclear protein; Zinc-finger.
SQ SEQUENCE 934 AA; 103010 MW; 814A43F847589B88 CRC64;
Query Match 5.9%; Score 311.5; DB 5; Length 934;
Best Local Similarity 18.6%; Pred. No. 3.2e-10;
Matches 195; Conservative 91; Mismatches 387; Indels 373; Gaps 34;
QY 48 LAHQNACSTDPYV-----VIIGQENPNNSASSSEPRBEHNNPQVWDTESNP 97
Db 1 KOHVSAASVPSVVTTPVTTGGTTITLGGPPLPKSEKEDKP-----PHGIEMKYNI 55
QY 98 PDSGSSVPTPTWGPERRGEESGHFLVAATGTAAGGGGLILASPKLGATPLPPESTPA 157
Db 56 EDLSQLEFTHVFG-----KIHGDVNNHQLAAHGGQL-----88
QY 158 PPPPPPPPPGVSGHLNPLILIEELRVLQQRQIHQOMTBOICROVLLLSLQGTGVA 217
Db 89 ---PPPPPLPPQVTS-HAASAAAAAASSTNNAAVVAWVASAAAAAASAGG--GL 142
QY 218 PASPSLPLPGTGTAASRKLPLRLPSIKRPVQTSKLASSSSSSSSSGATPPQAFHLVH 277
Db 143 PPATSS---GNGG-----OOVTVTTSSTSSGSGTSSGTTTAGELLMPK 185
QY 278 PLGSGHPESA---GGVGRSHKTPAP--SPALPDS-----TDOLIASPLAPP 320
Db 186 MGGHIGVGSNGSGNGGGGNNVALARPGETPATGTHVDCIGKMFQFRQLV--HRRVH 243
QY 321 STTGILAAQCLGARCLLEATASPEGLKPKKNGSGELSTGEVNGP-----L 364
Db 244 SRRKPFMCVCCGG---FTTSODILTR---HGKIHIS---GPMETCIQCFNVFANNTSL 292
QY 365 EKRPGHKK-----CRFCAKVPFSDSALOHLBSHTGRRPRKCNVCGNRFTRGNLKNVF 418
Db 293 ERMKRNHSHDPRPACTTCQKTFARKKHLDNHFRSHTGTPRCOYCATFTTKREHNVNV 352
QY 419 HNRKREYV---VQMPHVRPEHL--DYVITSSGLPYGMSVPRKAEBAEAPRGGVK 473
Db 353 RKNHTGETPRKIDCKSKSFYRKHENYVNIYMHHTGTPRHOCVCGKR-----YTRK 401
QY 474 PLVASTTALSTATESITLTLSTAGTATAPGLPAFNKKFVLMKAVEPRKNKADENTPPGSEGA 533

Db 402 EHLNHNMSHNET-----PRCELCGSKFSKREH----- 431
OY 534 ISGVAESSTATLQSLKMTSLPSMALLTNHF-K-STGSP--LPLCARALGASPSKSL 590
Db 432 -----FTNHLMTGETPHRCDCSKTF-----TRK 457
OY 591 QQLYEKIDRGAVAVTASASGAFTTAPAPSSSSSSGPNOCVICLRVLSCPRALRHYQ 650
Db 458 EHLNHN-RQ-----HTGESPHRCSYCKKTTTRKHLVNHMQ 494
OY 651 HGGEPFCKVCVGRAFSTGNTLRAHFVGHKASPARARONSCPICOKFTNAVTLQOHVM 710
Db 495 HTGEFPCYCTYCTKATRD-----HAWNHRQHTGESPHKCTYCTKTTTRKHLTNHVL 550
OY 711 HLGGOIPNGTALPEGGAANGSEOSTVSGASFPQOQSOQSPSEBELSEEEDEE 770
Db 551 HTG-----DSPHKCEY 561
OY 771 EEDVDTDSDSLANGRSGESGGERKALSVRGDSEBAGAEVEVGTVAATAAGKEMDSKETT 830
Db 562 COKTPTTRKHL-----NNHMR 577
OY 831 QQSLPPPPPSLDOPQPMEOGSSGVLGKKEGKPERSSSPASALPECEATSVTYE 890
Db 578 QHSS-DNPHCCVNCNKPFRKKEHLNHNMSCHTDGR-----FTCECTGSKSPKMG 627
OY 891 ELISLOBAMKPEGESSSKRACEVCGAFPSOALAEHOKTHPKESPLFTCYRCGFLER 950
Db 628 NLLFQRSHTKQCEMERPFACEKCPKNFTICGHLVSHMRSHSGEP-HACTLCSKAFAVER 686
OY 951 ATLKKHMLAH-----HOVQFPA 968
Db 687 GNLKRMKNNHPDAMPPPPVHPHPOTPAGVLTQVKQEVKPIITHHSATTTMTTIOQT 746
OY 969 PHGPONIALSLVPGCSPSITSTGLS 994
Db 747 AGAGGAGAVOLTRPGVPLVLTSTLIS 772
RESULT 38
Q9WAJ1 ID PRELIMINARY: PRT: 1893 AA.
AC 09WAJ1
DT 01-MAY-2000 (Tremblrel, 13, Created)
DT 01-MAY-2000 (Tremblrel, 13, Last sequence update)
DT 01-MAR-2002 (Tremblrel, 20, Last annotation update)
DE PEB protein (GH10905P).
GN PEB OR EG:66A1.1 OR CG_2212.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Artthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydrioidae; Drosophilidae; Drosophila.
OC NCBI_TaxId=7227;
XN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-BERKELEY;
RC MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amaralides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Mortan J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers J.-H.C., Blazey R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abriil J.F., Agayanni A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Bailew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borokova D., Botchan M.S., Bouck J., Brokstein P., Brotler P.,
RA Burlis K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cavley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos P., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrier S., Fleischmann W.,

RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.D., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibeagwa C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Kethum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spredling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskaas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster";
RL Science 287:2185-2195(2000).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-BERKELEY;
RA Stapleton M., Brokstein P., Hong L., Apabayani A., Carlson J.,
RA Champe M., Chavez C., Dorsett V., Farfan D., Fritse E., George R.,
RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
RA Nunoo J., Pacle J., Pargass V., Park S., Phouaneavong S., Wan K.,
RA Yu C., Lewis S.E., Rubin G.M., Celniker S.;
RL Submitted (OCR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AE003431; AAF45960.1;
DR EMBL: AY058335; AAL13564.1;
DR HSSP: P07248; ZADR.
DR Flybase: FBgn0003053; peb.
DR InterPro: IPR000822; znf_C2H2.
DR Pfam: PF00096; zf-C2H2; 12.
DR SMART: SM00355; znf_C2H2; 13.
DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; 11.
DR PROSITE: PS0157; ZINC_FINGER_C2H2_2; 10.
KW DNA-binding; Metal-binding; Zinc-finger.
SQ SEQUENCE 1893 AA: 205673 MW: 1B223EE96468A754 CRC64:

Query Match 5.8%; Score 308.5; DB 5; Length 1893;
Best Local Similarity 20.1%; Pred. No. 1,1e-09;
Matches 245; Conservative 132; Mismatches 426; Indels 413; Gaps 55;

OY 53 ACSTDPVWVYIIGQENPNSSASSEPPREGHN-NFQVMDTEHSNPPDSSSVFTDPTWG 111
Db 397 SCSTD-----VSSGSHSRSSSSSLNNNNNSHKANNLKDLE-----ELFVTEDDOTEN 446
OY 112 PERGGESSGFLV-----AATGTAAGGGGGLIILASPLGATPLRPES 154
Db 447 KORRLKTTINNNTIESEQOEDMDDEADADVAMLTSTPDVATLLGASASGA-----A 500
OY 155 TPAPPPPPPPP-----PGVSG-----HLNIPILIEELRVL-----QQR 190
Db 501 SRBPTRPSASPALISCPAGASPFETLPALCVHMDAMHSIDIPACRORDEVIFATIRQL 560
OY 191 QIHQMOMTEQICR-OVLLIGSLGQTVGAPASP-SELPETGTAASSTKPLPLPSPPIKPVQ 247
Db 561 QSH-----CCRLPNALAGLPLPILGASSSPLHNEPEDEEHGDDBDL-----E 603
OY 248 TSKTLASSSSSSSSSSSGAETPQAFPHLHYPLGSCQHPFSAAGVGRSHKPP----- 298
Db 604 QKERLASQSED-----FFHQLYLKHKTANGCAISHPSPITKEHPADTK 647
OY 299 -----APSPALPGSTDOLIASPHLPSTGTGLLAOCIGARGLAATASPGL- 345
Db 648 DLADIQSLNMTSSSSSFLRNFEQSVNTPNSSQYSLDG-----ROGEERAQDAFT 697
OY 346 -----LKPNGSGELSYGEVMGLEPKPGGRHKCRFCAKYFGS 382


```
698 SEFRMKLGEPPCKLCTAVFPLRLALKGNRVHIGAV-----GPAGPFCMNCPPYAVCD 752
383 DSAIQIHLRSHITGERPYKCNVCGNRFTRGNLKVHF-HRH-----REK----- 424
753 KAALVHRMRTNHGDRPEYECAYCNVAFYFTTKANCERHLNRNHGRTSREVKRAYVHPAEDA 812
425 -----YPHVQNNPHRPV-----EHLDVYITSSGLPYGMSVPEKA 459
813 GCEDSKSKLGEDLDATSFNSISPTPPRPVNESKSQLKHM--LIGENHLAPVNOQPLKI 870
460 E-----EEAATPGGVERKPLVASTTALSTATSLTLLSTAGTATAPGLPAFNK 508
871 QVKSILDOIYDKKPSAPAPQ0000000EKSAGALDFSMVDLDSKKP--TGCASLTPTAVTR 929
509 FVLKMAVEPKNADENTPPGSEGSATISGVAESSATLMO-----LSKIMTSLPSW 558
930 TTPPAVAVV-----TPGGVGTPTDLAAALEQ00LLAQ00LFGAGGEYMOQLFRSL-- 980
559 ALLTNHFKS--TGSFP--LPLCARALGASPSSETSKLOLVEKIDROGAVAVTSAASGAPYTS 616
981 -----MFGSQTSGPFPFFPMA-----PPPPQANPEKPRPVSPRRINIMPV--GVGVGVYV-- 1029
617 APAPSSSASSGNCVYICLRVLSCPRLRLHYG0HGERPFCVKVCGRAFSTRGNLRAHF 676
1030 -----PPGGVVKVYIKNGVL--MPKOKORRYRT--ERPFACEH0SARFTLRSMERHV 1078
677 -----VGH-----KASPARA----- 687
1079 KOHROFQAQORRSGIHVVRGRASNVAAAAAATAATYAGGPGSSGFGSNHHGH 1138
688 -----ONSCPICOKKFTNAVTLQ0HVRMH-----LGQILPNGTAL-----PE 725
1139 GHSGHSGHGHATISEQ--VKCALILAQ0LKAHKNTDILQ0ALAHGSSSVAGNPLHFGYPL 1196
726 GGGAQ0EKGSEBSTVSGASFPQ00SOOPS-----PEELSEEEEBEED 771
1197 TNPSPMHNGSSO---GNC0ATAMDDDEPKLIIIDENENHDEVEDVDDPDEDEDEE 1252
772 EEDVYTEDSL-----AGRGSSESGE--KAISVRGDSSEASGAEVEGVAATAAGKEM 823
1253 MDEPDEDELILDEQ0AKENAE0EBE0ELPKPLEJ0LTKREAQ0MAETI--LEQ0AIKAGKPL 1310
824 DSNKKT00SSLPPEPPPSLDOP--QPMEOGSS--GVLGKKEBCK--PERSSSP--- 873
1311 -SPPTKKNASPANPVATM0EPATITAPSTNPSLSLKTIM0AEVYKSLKEVASSEPKD 1369
874 -ASALTPEGE---ATSVT-----LVEELSLQE----- 896
1370 ESQDLVPYAKLVDMNATSQMNGFNSTFRPSDVANHMEQSDDEGLVASGSASESNNGTEDV 1429
897 ---AMRKEPGESS-----SRKACEVCG0AFPSQ0ALEHOKTTHKEGPLFTVCFCROGF 947
1430 TSSSSSEPKKSAVSLARNRVSCPYQ0RMFPMWSSSLRHLITHTGQ0K--FKCSHCILLF 1488
948 LERATLKHHMLLAHHOYQ 965
1489 TTKSNCDRHLLKRGHNV 1506
```

```
RESULT 39
077275 PRELIMINARY; PRT; 1891 AA.
077275;
01-NOV-1998 (Tremblrel. 08, Created)
01-NOV-1998 (Tremblrel. 08, last sequence update)
01-MAR-2002 (Tremblrel. 20, last annotation update)
EG:66A1.1 protein.
PEB OR EG:66A1.1 OR CG121212.
Drosophila melanogaster (Fruit fly).
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
```

```
OX NCBI_TaxID=7227;
RN [1]
RN SEQUENCE FROM N.A.
RA Ferraz C., Vidal S., Brun C., Bucheton A., Demaille J.G.;
RT Sequencing the distal X chromosome of Drosophila melanogaster.;
RL Submitted (Aug-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Benos P.;
RL Submitted (Oct-1998) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: NUCLEAR (By similarity).
DR EMBL; AL031227; CAA20227.1; -.
DR HSSP; P07248; 2ADR.
DR FlyBase; FBgn0003053; Peb.
DR InterPro; IPR000822; Znf_C2H2.
DR Pfam; PF00096; Zf-C2H2; 12.
DR PRINTS; FR00048; ZNCFINGER.
DR SMART; SM00355; Znf_C2H2; 13.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 11.
DR PROSITE; PS0157; ZINC_FINGER_C2H2_2; 10.
KW DNA-binding; Metal-binding; Nuclear protein; Zinc-finger.
SQ SEQUENCE 1891 AA; 205366 MW; 9E882364C36BB9BF CRC64;

Query Match 5.8%; Score 308; DB 5; Length 1891;
Best Local Similarity 20.3%; Pred. No. 1,2e-09;
Matches 247; Conservative 130; Mismatches 425; Indels 416; Gaps 56;

OY 53 ACSTDPYMWITIGQENPNSSASSEPREEGHN-NPOVMDTEHSNPDSGSSVPTDPTWG 111
DB 397 SCSTD-----VSSGSHSKSSSSSLNNNNNSHKANNLIKDE-----ELEVEATEDOTEN 446
OY 112 PERGEESGHFLV-----AATGTAGGGGGLILASPKIGATPLPPPS 154
DB 447 KORRLKTTINNIIIESEQEDMDDEADADADVAMLTSTPOVATILLAGASASG-----A 500
OY 155 TPAPPPPPPPPP-----PGVSG-----HNTPLLEELRYL---QQR 190
DB 501 SRSPTPSPASAPALLSCPACGASDEFETPLALCVNLDAMSHSDIPKCHDCVFIATNRDL 560
OY 191 QIHOMQWEOICR-OVLLIGSLGQVFGAPASP--SELPGTASTASTKPLPLPSPIKPYQ 247
DB 561 QSH-----CRLNMLALAGLPRLLAGASSPLINEPEDEH0DDEL-----E 603
OY 248 TSKTILASSSSSSSSGAEPTKQAFPHLYHPLGSQHPFSAGVGRSHKPTP----- 298
DB 604 OKERLASQSED-----FFHOLYLKHKTANCGAISHPPSPKHNPRADTK 647
OY 299 -----APSPALPGSTDLITLAPPLAPPTTGILAAQ0CGAARGLEATAPSPGL- 345
DB 648 DLADIQSTILNMTSSSSSFLRNEQ0SVNTPNSSQYSLDG-----RDQEEAD0AFT 697
OY 346 -----LKPKN0SGELSYGEVMGPLEKPGGRHKCFKCAKVFPS 382
DB 698 SEFRMKLGEPPCKLCTAVFPLRLALKGNRVHIGAV-----GPAGPFCMNCPPYAVCD 752
OY 383 DSAIQIHLRSHITGERPYKCNVCGNRFTRGNLKVHF-HRH-----REK----- 424
DB 753 KAALVHRMRTNHGDRPEYECAYCNVAFYFTTKANCERHLNRNHGRTSREVKRAYVHPAEDA 812
OY 425 -----YPHVQNNPHRPV-----EHLDVYITSSGLPYGMSVPEKA 459
DB 813 GCEDSKSKLGEDLDATSFNSISPTPPRPVNESKSQLKHM--LIGENHLAPVNOQPLKI 870
OY 460 EEAATPGGVERKPLV-----ASTTALS-ATESLTLTSTAGTATAPGLPAF 506
DB 871 QVKSILDIQ--LVDKKPSAPAPQ0000000EKSAGALDFSMVDLDSKKPPTGAS----- 921
OY 507 NKFLVLMKAVEPKNADENTPPGSEGSATISGVAESSATLMO-----LSKIMTSL 555
DB 922 ---LTPAVPTTPPAVAVPYTPGGVGTPTDLAAALEQ00LLAQ00LFGAGGEYMOQLFRSL 977
OY 556 PSMALLTNHFKS--TGSFP--LPLCARALGASPSSETSKLOLVEKIDROGAVAVTSAASGAP 613
```


Db	978	-----MFOSTSGFPFFPMA---PPPOANEKPPMVSPPRINMPV-GVGVGVP	1025
QY	614	TTSAAPSSSSASGPNOCVTCILRVLSCPRALRLHYGSGEERPFCKYCGRASTGNLR	673
Db	1026	V-----PBGCPKMYNKNVNL-MPKOKRRRT---EPPFACEHCSARTLASMME	1072
QY	674	AHF-----VGH-----KASPARA-----	687
Db	1073	RHVQKHPOFYAORORSGHHVMRGASNVAAAAAATAVMAGPGSGGSGSNH	1132
QY	688	-----QNSCRPCO---KFTNAVTLOQHR-----MHGGO	715
Db	1133	HGHGHSCHSGHAPRISEYKCALIAOOLKAHKNTDLLOALAHGSSVAGNPLHFGYP	1192
QY	716	IPNGGTALPEGGAOENGSEOSTVSGASF---POOQSQPSPEELSEEEDEDEE	771
Db	1193	LTNPS---FMHNSSQNGGATATADDEPKLIDEDENHHEVEADVDDEDEDEE	1249
QY	772	EEDVTDDSL-----AGRGSESGE---KAISVRGDSSEASGAEEVGTVAATAAGKEM	823
Db	1250	MEDEDEPELILDEOPAEKEAEEBOELPKPLEQGTKEAOKMAETI--LEQATKAGKPL	1307
QY	824	DSNEKTOQSSLRPPRPDLDOP---QPMEOGSS--GYLGGKEEGK--PERSSSP---	873
Db	1308	-SPPTKEANSPANPTVATTMOEPATITAPSTNPSLTKMTIAQAEYVKGSLKEVASSPKFD	1366
QY	874	-ASALTPEGE---ATSVT-----LVELLSIQE-----	896
Db	1367	ESQDILVPYAKLVNATNSQNNGFNSYFRPSDVANHMEQSDPEELVAGSGSSESNNSGEDV	1426
QY	897	---AMKEPGESS-----SRKACEVCGQAFRPSQALAEHOKTHPREGPLFTVCVFCROGF	947
Db	1427	TSSSSSEPKKKAAYSLAPRNVSCPYCQRMFPMSSSLRHILTHTGQKP-FKCSHCLLF	1485
QY	948	LERATLKHHMLLHHQYQ	965
Db	1486	TTKSNCDRHLRLKHGNYE	1503
RESULT 40			
ID	090783	PRELIMINARY;	PRT: 1173 AA.
AC	090783;		
DT	01-NOV-1996	(Tremblrel. 01, Created)	
DT	01-NOV-1996	(Tremblrel. 01, Last sequence update)	
DT	01-MAR-2002	(Tremblrel. 20, Last annotation update)	
DE	Prock12 (Fragment).		
OS	Gallus gallus (Chicken).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;		
OC	Gallus.		
OX	NCBI_TaxID=9031;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=BONE MARROW CELLS;		
RA	MEDLINE=95047430; PubMed=7958949;		
RA	Schuetz B., Niessing J.,		
RT	"Cloning and structure of a chicken zinc finger cDNA:restricted		
RT	expression in developing neural crest cells.";		
RL	Gene 148:227-284(1994).		
DR	EMBL; X56805; CAA40140.1; .		
DR	HSSP; P08047; 1SP2.		
DR	InterPro: IPR000822; znf_C2H2.		
DR	Pfam: PF00966; zf_C2H2; 16.		
DR	SMART; SM00355; znf_C2H2; 16.		
DR	PROSITE; PS00028; ZINC_FINGER_C2H2_1; 15.		
DR	PROSITE; PS00157; ZINC_FINGER_C2H2_2; 15.		
KW	DNA-binding; Metal-binding; Zinc-finger.		
FT	NON_TER		
FT	SEQUENCE	1173 AA; 127327 MW; 493BF4C07300FA10 CRC64;	
Query Match			
5.8%; Score 307; DB 13; Length 1173;			

Best Local Similarity 21.4%; Pred. No. 7.6e-10;			
Matches 189; Conservative 90; Mismatches 314; Indels 292; Gaps 36;			
QY	292	RSKHPTAPSPALPGSDQLASPHLAPSTTGLAOLGAARGLAATSPGLKKNK	351
Db	9	RPYKSECPKAFKRSLSLIQHSVAH-----TGLRFKC--ALGIAF-----KW	50
QY	352	SGELSYGEVMGPLEKPGGRHNCRCFAKYGSDSALOIHRSHTGERPYKNCVGNREPTR	411
Db	51	SSHXYHLRQHTGERP---YKCTSCPFAFNSSLRHHNHHTGERPYVSACGAKTQS	107
QY	412	GNLKVHNRHREKYRHYQMNPRVPEHLDYITSSG--LPYGNVPREKAEERATPGG	469
Db	108	TNLRQHTHTGERPY-----ACSHCSKPTTHSNLLHQRTHSSTRSHKCPACPAF	160
QY	470	VERKPLVASTALSTATSITLSTAGTATAGLPAFNKFVLMKAPRKAADENT----	525
Db	161	VSDACLQKHLQSHNASPLLPSPLSQLSPP-----LLEAVEMLYKTECTYLSLP	212
QY	526	-----PPGSGSAISGVAESSTATLMOLSKLMTSLPSMAALTNHEKSTGS--	570
Db	213	COBQLGHQSHRQPATPGD-VTATATHTRCPTCGKTFKNSSG-----LARHRSHGAER	265
QY	571	-FPLPLCARALGASPSSETSKLQOLYEKIDROGAVAVTSAASGAPTSAPAPSSASSGPN	629
Db	266	RYKSCQCHRSFG-----QLAGLGHORG---HSAETPHRPATPTPTSVPSERPY	312
QY	630	QCVIY-----	642
Db	313	QCTECGKAFKSGSLRYHMRDHTGERPYKSCPCPAFKRSSLLAHQVHTGLAYKCP	372
QY	643	A-----LRHYGO-----HGERRPFCKYCGRA	665
Db	373	CGLTFKSSNHQYHLRLHTGERPYRCPCPKAFKNTSCGLPHQLHTGERPHNACPIGKA	432
QY	666	FSTRGNLRAH--FVG-----HKASPARAONSCPICOKKFTNA	701
Db	433	FTQTSNLROHRTHTGERPYACSHCGKFTTHSSNLQLRQTHSSABPHQCLCPKAFVMA	492
QY	702	VTLQAHYRMHLGGQIPNGTALPEGGAOENG-----SEQSYSGASGSPQOQSQ-	753
Db	493	SYLQRHLRTNAAG--PKGS---PRPALTPRODGPVYLQALSLLEVTPADAHFTLLQTPQG	547
QY	754	-----PSPEE--ELSEEEEEDEEDEDVTDDESL--AGRGS-----	786
Db	548	LQILPSPPAPQKLLLPATQPRPKHGESPTPGQSLLVPSTGTLPLRLQAVTAVP	607
QY	787	ESGGEKATSVRGDSEE--ASGAEEVGTVAATAAGKEDSNKKT--TOQSSLRPPPPPD	842
Db	608	QGTGTGLIVLQGLPEQRLHPAGIRPGQAAVEGAAY--RLQANEVTVNQALQALPQPDVT	664
QY	843	SLD-----QPPRMQSGSGVLGKKEGGKPERSSSPASALTPREGENTSVTLVEELS	893
Db	665	SIPLOATDMTNVQALQALPQ-----PTDVTINIQALQATEVTNVQ---	701
QY	894	LOEAMRKEPGESSSRKACEVCGQAFRPSQALAEHOKTHPREGPLFTVCVFCROGLERATL	953
Db	702	LQALLQPTDVTNIOQATEVANVL--QALSQPDVTNIE-----LQAAEV	745
QY	954	KKHMLLHHQVQPARPHQPNITIALSLVPGCSPTSTGSLSPPR	998
Db	746	TN-----AHLQVLP---QPSVETNIQLODG---EVTNVQQLQTLPQ	779

Search completed: January 13, 2003, 15:22:30
Job time : 73.0643 secs

Run on: January 13, 2003, 15:17:40 ; Search time 37.9432 Seconds

(without alignments)
3518.868 Million cell updates/sec

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

Title: US-09-988-117-3

Sequence: 1 MAOETGSSSRLLGPGCEPAE.....SSIPSPGLSPFPKKDDPTMP 1002

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database : A_Geneseq.101002:*

1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	5301	100.0	1002	23	AAE17955	Mouse Sal2 protein
2	4677.5	88.2	1005	23	AAE17954	Human Sal2 protein
3	4672.5	88.1	1005	23	AAE17967	Human Sal2 protein
4	4669.5	88.1	1005	23	AAE17968	Human Sal2 protein
5	4625.5	87.3	1019	22	AAW79822	Human protein SBO
6	4615.5	87.1	1007	22	AAW78838	Human protein SBO
7	845.5	15.9	813	22	AAW91193	Human protein sequ
8	765	14.4	1373	22	ABB62368	Drosophila melanog
9	736.5	13.9	1267	22	ABB61050	Drosophila melanog
10	697	13.1	172	22	AAO01962	Human polypeptide

11	529	10.0	200	22	AAB92796	Human protein sequ
12	482.5	9.1	330	22	ABB30894	Peptide #3545 enco
13	482.5	9.1	330	22	ABB36076	Peptide #3582 enco
14	482.5	9.1	330	22	AAW56858	Human brain expres
15	482.5	9.1	330	22	AAW69245	Human bone marrow
16	482.5	9.1	330	22	AAW17074	Peptide #3508 enco
17	482.5	9.1	330	22	AAW29568	Peptide #3605 enco
18	482.5	9.1	330	22	AAW04775	Peptide #3457 enco
19	482.5	9.1	330	22	ABG38855	Human peptide enco
20	482.5	8.3	336	22	ABG32272	Peptide #4923 enco
21	438.5	8.3	336	22	ABG37530	Peptide #5036 enco
22	438.5	8.3	336	22	AAW58186	Human brain expres
23	438.5	8.3	336	22	AAW70641	Human bone marrow
24	438.5	8.3	336	22	AAW18485	Peptide #4919 enco
25	438.5	8.3	336	22	AAW30956	Peptide #4993 enco
26	438.5	8.3	336	22	AAW06075	Peptide #4757 enco
27	438.5	8.3	336	22	ABG40328	Human peptide enco
28	337	6.4	927	22	AAW80283	Human protein SBO
29	330.5	6.2	619	22	AAW95103	Human protein sequ
30	324.5	6.1	869	22	AAW79299	Human zinc finger
31	322.5	6.1	702	23	ABW04333	Human protein sequ
32	321.5	6.0	498	22	AAW94658	Human nucleic acid
33	320.5	5.9	799	21	AAW21003	Drosophila melanog
34	311.5	5.9	744	22	ABW59449	Drosophila melanog
35	311.5	5.9	744	22	ABW66002	Drosophila melanog
36	311.5	5.9	744	22	ABW66572	Drosophila melanog
37	309	5.8	934	22	ABW69958	Drosophila melanog
38	309	5.8	1893	22	ABW59829	Novel human diagno
39	305.5	5.8	1196	22	ABG28743	Human protein sequ
40	304	5.7	725	22	AAW95116	Human protein sequ
41	303	5.7	722	22	AAW39029	Human polypeptide
42	300.5	5.7	833	22	AAW99340	Human Rlt1 beta pr
43	299	5.6	894	22	AAW99335	Human Rlt1 protein
44	297.5	5.6	515	22	ABW31309	Peptide #3960 enco
45	297.5	5.6	515	22	ABW36515	Peptide #4021 enco
46	297.5	5.6	515	22	AAW21854	protein #3853 enco
47	297.5	5.6	515	22	AAW57278	Human Dratn expres
48	297.5	5.6	515	22	AAW69683	Human bone marrow
49	297.5	5.6	515	22	AAW17495	Peptide #3929 enco
50	297.5	5.6	515	22	AAW30016	Peptide #4053 enco
51	297.5	5.6	515	22	AAW05170	Peptide #3852 enco
52	297.5	5.6	515	22	ABG39299	Human peptide enco
53	297	5.5	890	22	AAW99339	Human Rlt1 alpha'
54	292.5	5.5	812	22	AAW99337	Mouse Rlt1 beta pr
55	292	5.5	624	23	ABW57324	Mouse ischemic co
56	290	5.5	1185	20	AAW3497	Novel human diagno
57	289	5.5	877	22	ABG16854	Human atrophin I p
58	289.5	5.5	567	22	AAW3658	Novel human diagno
59	288.5	5.4	880	22	AAW99336	Human polypeptide
60	288.5	5.4	884	22	AAW99334	Mouse Rlt1 alpha'
61	288.5	5.4	2639	22	ABG15016	Mouse combined DNA
62	283	5.3	567	21	AAW93988	Novel human diagno
63	282.5	5.3	882	22	ABG06465	Novel human diagno
64	281	5.3	1212	22	ABG00399	Novel human diagno
65	281	5.3	1472	22	AAU31578	Novel human secreto
66	279.5	5.3	491	21	AAW52153	Human secreted pro
67	278.5	5.3	1230	22	AAU30831	Novel human secreto
68	278.5	5.3	582	23	ABW7429	Novel human diagno
69	277.5	5.2	832	22	ABG10543	Novel human diagno
70	277.5	5.2	2703	22	ABW63299	Drosophila melanog
71	277.5	5.2	675	22	AAW94388	Human protein sequ
72	276.5	5.2	556	22	ABW60575	Drosophila melanog
73	273.5	5.2	701	22	AAW99341	Human Rlt1 gamma p
74	272.5	5.1	577	22	AAU15930	Human novel secret
75	270.5	5.1	839	21	ABW43281	Human OREFX OREF3045
76	270.5	5.1	756	22	ABG08429	Novel human diagno
77	269.5	5.1	839	22	AAW93280	Human protein sequ
78	269.5	5.1	1520	22	ABW27130	Novel human diagno
79	269.5	5.1	691	22	AAW99338	Mouse Rlt1 gamma p
80	269	5.1	810	21	AAW73351	HTM clone 1484257
81	269	5.1	1054	22	ABW57993	Drosophila melanog
82	269	5.1	714	22	ABG17368	Novel human diagno
83	268.5	5.1				

84	268.5	5.1	1252	22	AAAT9739	Human protein SEO
85	268	5.1	1205	22	ABG05068	Novel human diagno
86	268	5.1	1205	22	ABG15312	Novel human diagno
87	268	5.1	1214	22	ABG27121	Novel human diagno
88	267.5	5.0	1274	22	ABG37532	Novel human protei
89	267.5	5.0	820	22	ABB45565	Drosophila melanog
90	267	5.0	613	22	ABB95862	Human protein sequ
91	266.5	5.0	1083	22	ABB12422	Human bone marrow
92	266	5.0	481	22	ABB92770	Human protein sequ
93	266	5.0	481	23	ABB97545	Novel human protei
94	266	5.0	791	22	ABB67268	Drosophila melanog
95	266	5.0	845	22	ABB64280	Drosophila melanog
96	265	5.0	755	22	AAA40916	Human polypeptide
97	265	5.0	809	22	AAA38689	Human polypeptide
98	264	5.0	804	22	AAA41058	Human polypeptide
99	263.5	5.0	529	22	AAA93633	Human polypeptide
100	263	5.0	803	22	AAA92722	Human polypeptide

ALIGNMENTS

RESULT 1

AAE17955 ID AAE17955 standard; Protein; 1002 AA.

AC AAE17955;

XX 07-MAY-2002 (first entry)

XX Mouse Sal2 protein.

XX Mouse; proliferative disorder; tumor host range mutant virus; cancer;
T-HR mutant; Sal2 protein; ovarian tumour.

XX Mus musculus.

XX Key Location/Qualifiers

XX Misc-difference 252 /note= "Encoded by CTG"

XX WO200204596-A2.

XX 17-JAN-2002.

XX 05-JUL-2001; 2001WO-US21354.

XX 07-JUL-2000; 2000US-216723P.

XX 19-MAR-2001; 2001US-0812471.

XX 19-MAR-2001; 2001US-0812633.

XX (HARD) HARVARD COLLEGE.

XX (BGM) BRIGHAM & WOMENS HOSPITAL INC.

XX Benjamin TL, Ll D, Mok SC, Cramer DW, Ma Y;

XX WPI: 2002-164637/21.

XX N-PSDB; AAD28652.

XX Detecting protein involved in susceptibility to proliferative disease,
by infecting normal and abnormal proliferating cells with mutant virus,
detecting mutated protein allowing growth of mutant on abnormal cells

XX -

XX Disclosure; Page 85-87; 92pp; English.

XX The invention relates to a method for the identification of genes
C and their encoded proteins involved in susceptibility to proliferative
C disorders, including cancer using a tumor host range mutant (T-HR
C mutant) virus. The invention also provides the use of Sal2 genes and
C proteins in methods of identifying a mammal having, or at a risk of
C acquiring a proliferative disease. T-HR mutants are used to kill cancer
C cells such as one carrying a Sal2 alteration. Transgenic and knockout

CC mouse comprising Sal2 nucleic acid are useful as research tools to
CC determine genetic and physiological features of cancer and for
CC identifying compounds that can affect ovarian and other tumours. The
CC present sequence is mouse Sal2 protein.

XX Sequence 1002 AA:

Query Match 100.0%; Score 5301; DB 23; Length 1002;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1002; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAQETGSSRLGGCGCPAERGSDAEEHHPOVAKCCAQFSPTEFLAHQNSCCTPPV 60
Db 1 MAQETGSSRLGGCGCPAERGSDAEEHHPOVAKCCAQFSDPTEFLAHQNSCCTDPV 60
QY 61 MVIIGQENPSNSASSAPRDEGHSRQVMDTEHNPDPGSSGPPDPMPGPRRGESS 120
Db 61 MVIIGQENPSNSASSAPRDEGHSRQVMDTEHNPDPGSSGPPDPMPGPRRGESS 120
QY 121 GQFLVATGTAAAGGGGLILASPRLGATPLPESTPAPPPPPPPPPPGVSGHLNPLT 180
Db 121 GQFLVATGTAAAGGGGLILASPRLGATPLPESTPAPPPPPPPPPPGVSGHLNPLT 180
QY 181 LEBELRVLQQRQIHOMQMTQICRQVLLLSGQTFGAPSPSELPTGAASSTKPLPLF 240
Db 181 LEBELRVLQQRQIHOMQMTQICRQVLLLSGQTFGAPSPSELPTGAASSTKPLPLF 240
QY 241 SPKPAQTGKTTASSSSSSSSGAEPKQAFPHLYHPLGSGHPFSVGVGRSHKTPAPS 300
Db 241 SPKPAQTGKTTASSSSSSSSGAEPKQAFPHLYHPLGSGHPFSVGVGRSHKTPAPS 300
QY 301 PALPGSTDOLIASPHLAFPGTTGLLAOCIGAARGLAAASPGLLKPKNGSELGGEVY 360
Db 301 PALPGSTDOLIASPHLAFPGTTGLLAOCIGAARGLAAASPGLLKPKNGSELGGEVY 360
QY 361 SSIKPGGRHRCRCFAKVFSDSALQIHLSHTGERPKVCNVCNRRFTRGMLKVPHNRH 420
Db 361 SSIKPGGRHRCRCFAKVFSDSALQIHLSHTGERPKVCNVCNRRFTRGMLKVPHNRH 420
QY 421 RKXYHVQNNPVPVPHLDVITSSGLPYGMSVPEKKEEAGTPGGGVERRPLVASTTA 480
Db 421 RKXYHVQNNPVPVPHLDVITSSGLPYGMSVPEKKEEAGTPGGGVERRPLVASTTA 480
QY 481 LSATESILTLSTGTSTAVALPGLPTFNKFLMKAVEPKSKABENPPGSGSIAAGVDSG 540
Db 481 LSATESILTLSTGTSTAVALPGLPTFNKFLMKAVEPKSKABENPPGSGSIAAGVDSG 540
QY 541 SATRMQLSKLVTSIPSWALLTNHLKSTGSEFPFVYLEPLGASPSSTSKLQQLVEKIDRQG 600
Db 541 SATRMQLSKLVTSIPSWALLTNHLKSTGSEFPFVYLEPLGASPSSTSKLQQLVEKIDRQG 600
QY 601 AVAVASTAGAPTTSPAPSSSAGPNOCTICLRYLSCPRALRLHYGHGGERPKCKYC 660
Db 601 AVAVASTAGAPTTSPAPSSSAGPNOCTICLRYLSCPRALRLHYGHGGERPKCKYC 660
QY 661 GRAFSTRGTLRAHFVGHKTSPPARAQNSCPIQCKKFTNAVYLQOHVNRHLGGQINGGSA 720
Db 661 GRAFSTRGTLRAHFVGHKTSPPARAQNSCPIQCKKFTNAVYLQOHVNRHLGGQINGGSA 720
QY 721 LSEGGAQOENSSEOSTASGPGSPPOQSQPDEEMSEEEDEEDVTDSDSLAG 780
Db 721 LSEGGAQOENSSEOSTASGPGSPPOQSQPDEEMSEEEDEEDVTDSDSLAG 780
QY 781 RGSSEGGKATSVRGDSSEFVSGAEEVATVVAAPTYKEDSNKAKOHTLPPPPPPDN 840
Db 781 RGSSEGGKATSVRGDSSEFVSGAEEVATVVAAPTYKEDSNKAKOHTLPPPPPPDN 840
QY 841 LDHQPMDQGTSDVSCAMEEAKLEGISPMALTOGEGSTPLVBEINLPEAMKRPD 900
Db 841 LDHQPMDQGTSDVSCAMEEAKLEGISPMALTOGEGSTPLVBEINLPEAMKRPD 900
QY 901 ESSRRKACEVCGSPPTOTALAEHQTHPKDGLFTVCRCRQFLDRATLKKHMLAHQ 960
Db 901 ESSRRKACEVCGSPPTOTALAEHQTHPKDGLFTVCRCRQFLDRATLKKHMLAHQ 960

Db 901 ESSGRKACEVCGSPTPTQTALEHOKTHRKDGLFTVCVCRGFLDRATLTKHMLAHQ 960
Oy 961 VPPAPHPQNIATLTLVPGCCSSIPSPGLSPPRKDDPTWP 1002
PT |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
PT 961 VPPAPHPQNIATLTLVPGCCSSIPSPGLSPPRKDDPTWP 1002
PT
RESULT 2
ID AAE17954 standard: Protein: 1005 AA.
XX AAE17954:
AC AAE17954:
DT 07-MAY-2002 (first entry)
XX Human Sal2 protein.
XX DE
XX Human, proliferative disorder: tumor host range mutant virus: cancer;
KM T-HR mutant; Sal2 protein; ovarian tumour; chromosome 14q12.
XX Homo sapiens.
OS
XX
FH Location/Qualifiers
FT Misc-difference 15 /note= "Encoded by TGC of the inverse complementary strand of the sequence shown as SEQ ID NO:2 (AAD28651)"
FT Misc-difference 22..23 /note= "Encoded by bases 4611-5022 of the inverse complementary strand of the sequence shown as SEQ ID NO:2 (AAD28651)"
FT Misc-difference 120 /note= "Encoded by CCA of the inverse complementary strand of the sequence shown as SEQ ID NO:2 (AAD28651)"
FT Misc-difference 545 /note= "Encoded by CGC of the inverse complementary strand of the sequence shown as SEQ ID NO:2 (AAD28651)"
FT Misc-difference 552 /note= "Encoded by GTG of the inverse complementary strand of the sequence shown as SEQ ID NO:2 (AAD28651)"
FT Misc-difference 573 /note= "Encoded by TTC of the inverse complementary strand of the sequence shown as SEQ ID NO:2 (AAD28651)"
FT Misc-difference 575 /note= "Encoded by TAT of the inverse complementary strand of the sequence shown as SEQ ID NO:2 (AAD28651)"
FT Misc-difference 576 /note= "Encoded by GTG of the inverse complementary strand of the sequence shown as SEQ ID NO:2 (AAD28651)"
FT Misc-difference 577 /note= "Encoded by ATC of the inverse complementary strand of the sequence shown as SEQ ID NO:2 (AAD28651)"
FT Misc-difference 578 /note= "Encoded by GAG of the inverse complementary strand of the sequence shown as SEQ ID NO:2 (AAD28651)"
FT Misc-difference 579 /note= "Encoded by CCC of the inverse complementary strand of the sequence shown as SEQ ID NO:2 (AAD28651)"
FT
XX WO200204596-A2.
XX
XX 17-JAN-2002.
XX
XX 05-JUL-2001: 2001MO-U\$21354.
XX
XX 07-JUL-2000: 2000US-216723P.
XX 19-MAR-2001: 2001US-0812471.
XX 19-MAR-2001: 2001US-0812633.
XX
XX (HARD) HARVARD COLLEGE.
XX (BGHM) BRIGHAM & WOMENS HOSPITAL INC.
XX Benjamin TL, LI D, Mok SC, Cramer DW, Ma Y;
XX

DR WPI: 2002-164637/21.
DR N-PSDB: AAD28651.
XX
XX
PT Detecting protein involved in susceptibility to proliferative disease,
PT by infecting normal and abnormal proliferating cells with mutant virus,
PT detecting mutated protein allowing growth of mutant on abnormal cells
PT
XX
PS Disclosure: Page 77-79; 92pp: English.
XX
XX The invention relates to a method for the identification of genes
CC and their encoded proteins involved in susceptibility to proliferative
CC disorders, including cancer using a tumor host range mutant (T-HR
CC mutant) virus. The invention also provides the use of Sal2 genes and
CC proteins in methods of identifying a mammal having, or at a risk of
CC acquiring a proliferative disease. T-HR mutants are used to kill cancer
CC cells such as one carrying a Sal2 alteration. Transgenic and knockout
CC mouse comprising Sal2 nucleic acid are useful as research tools to
CC determine genetic and physiological features of cancer and for
CC identifying compounds that can affect ovarian and other tumours. The
CC present sequence is human Sal2 protein. The Sal2 gene is located on
CC chromosome 14q12.
XX
SQ Sequence 1005 AA:
Query Match 89.2%; Score 4677.5; DB 23; Length 1005;
Best Local Similarity 89.0%; Pred. No. 2.2e-284;
Matches 894; Conservative 25; Mismatches 83; Indels 3; Gaps 3;
Oy 1 MAOETSSSRILGPGCGPAERGDASEBHPQVCAKCAQSFDPTEFLAHONSCCTDPV 60
DB 1 MAHESRRSRILGVPAGPAELGDAEEDHPQVCAKCAQFDPTEFLAHONSCCTDPV 60
Oy 61 MYTIGQENPNSNSASAPREHRSRQVMDTEHNPDSGSGSPDDPMGPBRGESS 120
DB 61 MYTIGQENPNSNSASAPREHRSRQVMDTEHNPDSGSGSPDDPMGPBRGESS 120
Oy 121 GQFLVATGTAAGGGGGLILASPKIGATPLPESTPAPPPEPPPPPGVSGHLNPLI 180
DB 121 GHLVATGTAAGGGGGLILASPKIGATPLPESTPAPPPEPPPPPGVSGHLNPLI 180
Oy 181 LELRLVQQRQIHOMOMTEOTICROVLLGSLQYTGAPASPSLPTGTGAASRKLPLF 240
DB 181 LELRLVQQRQIHOMOMTEOTICROVLLGSLQYTGAPASPSLPTGTGAASRKLPLF 240
Oy 241 SPIKPAQTGTTA-SSSSSSSSSGAEPKQAFPHLYHPIGSOHFFVGVGGRSHKPTAP 299
DB 241 SPIKPAQTGTTA-SSSSSSSSSGAEPKQAFPHLYHPIGSOHFFVGVGGRSHKPTAP 299
Oy 300 SPALPGSTDLIASPHLAFPGTTGLAAQCLGAARGLAASAPGLKPKNGSGELGYGEV 359
DB 301 SPALPGSTDLIASPHLAFPGTTGLAAQCLGAARGLAASAPGLKPKNGSGELGYGEV 360
Oy 360 ISSLEKRGRRHKRCFAKVFSGDSALOHLRSHGTGRPKCNVCGNFTTGRNKKVPHR 419
DB 361 MGPLEKPGGRHKRCFAKVFSGDSALOHLRSHGTGRPKCNVCGNFTTGRNKKVPHR 420
Oy 420 HREKYPHYOMNPHVPEHLDVYITSSGLPYGMSVPEPEKAEPEATPGGVERKPLVASTT 479
DB 421 HREKYPHYOMNPHVPEHLDVYITSSGLPYGMSVPEPEKAEPEATPGGVERKPLVASTT 480
Oy 480 ALSATESLTLLSTGTAVAAPGLPTFNKTVLMAKAVEPKSKADENTPPGSECSAIVADS 539
DB 481 ALSATESLTLLSTGTAVAAPGLPTFNKTVLMAKAVEPKSKADENTPPGSECSAIVADS 540
Oy 540 GSATRMQLSKLVTSIPSWALLTNHLKSTGSPRPYVLEPLGASPSERSKIQQLVEXIDRQ 599
DB 541 STATIMQLSKLMTSLPWWALLTNHRSSTGSPFLCARALGASSETSKIQQLVEXIDRQ 600
Oy 600 GAVAVASTASGAPTSAPAPSSSA-SGPNCVCYICLRVLSGPRALRLHYGQHGGERPFCK 658
DB 601 GAVAVASTASGAPTSAPAPSSSA-SGPNCVCYICLRVLSGPRALRLHYGQHGGERPFCK 660

Y	659	VCRRAEFTSTGNLRARHVFVGHKTSAPAAARQNSCPICQKKFTNAVTLQOHVRLHIGQIPNGG	718
Y	661	VCGRASFSTGNLRARHVFVGHKASPAARAQNSCPICQKKFTNAVTLQOHVRLHIGQIPNGG	720
Y	719	SALSCGCGAQAQENSSQOSTASGSGSEFPQPSQOPSPPEEMSS-EEEEDEDEEEDVTDSDS	777
Y	721	TALPEGCGAQAQENSSQOSTASGSGSEFPQPSQOPSPPEELSEEEEDDEEEDVTDSDS	780
Y	778	LAGRSESGEGEALISVRGSESEVSGAEEVATSVAAPTVYKEMDSNEKAPHTLPPPPP	837
Y	781	LAGRSESGEGEALISVRGSESEVSGAEEVATSVAAATAGKEMDSNEKXTQOOSLPPPPP	840
Y	838	PDNLDPQPMEGCTSDVSCGAMBEAKLEGTSPMAALTOGEGNSTPLVEELNLPKMK	897
Y	841	PDSLDQPPQPMEGSSVGLGKEGGKPERSSSPASALTPREGATSVIIVELSLQEAAMK	900
Y	898	DPESSGKACAEVCGSGSFPTQTALAEHOKTHPRDGLFTVCVCRQGFRLRATLKKHMLA	957
Y	901	EPGESSSKACAEVCGGAFPSQALAEHOKTHPRKEGRLFTVCVCRQGFRLRATLKKHMLA	960
Y	958	HHQVFPAPHGPNATATSLVPGCSSIPSGISPPRKDDPTMP	1002
Y	961	HHQVFPAPHGPNATATSLVPGCSSITSTGLSPPRKDDPTIP	1005
RESULT 3			
D	AAE17967	standard; Protein; 1005 AA.	
X	AAE17967;		
X	07-MAY-2002	(first entry)	
X		Human Sal2 protein mutant (S73C).	
X		Human; proliferative disorder; tumor host range mutant virus; cancer;	
X		T-HR mutant; Sal2 protein; ovarian tumour; mutant; mutlein.	
X		Homo sapiens.	
S		Synthetic.	
X	Key	Location/Qualifiers	
T	Misc-difference 73	/note= "Wild-type Ser substituted with Cys"	
T	WC0200204596-A2.		
X	17-JAN-2002.		
X	05-JUL-2001; 2001WO-US21354.		
F	07-JUL-2000; 2000US-216723P.		
R	19-MAR-2001; 2001US-0812471.		
T	19-MAR-2001; 2001US-0812633.		
X	(HARD) HARVARD COLLEGE.		
A	(BGHM) BRIGHAM & WOMENS HOSPITAL INC.		
X	Benjamin TL, Li D, Mok SC, Cramer DW, Ma Y;		
X	WPI; 2002-164637/21.		
X			
X			
T	Detecting protein involved in susceptibility to proliferative disease.		
T	by infecting normal and abnormal proliferating cells with mutant virus		
T	detecting mutated protein allowing growth of mutant on abnormal cells		
T	-		
X	Claim 35; Page -; 92pp; English.		
X			
X	The invention relates to a method for the identification of genes		
X	and their encoded proteins involved in susceptibility to proliferative		
X	disorders, including cancer using a tumor host range mutant (T-HR		
X	mutant) virus. The invention also provides the use of Sal2 genes and		

Query Match	Best Local Similarity	Score	DB	Length
Matches 893;	Conservative 25;	Mismatches 84;	Indels 3;	Gaps 3;
Sequence 1005 AA:				
1 MAEETGSSRLRGCGGPEAERGGASSEHHNPVCAKCAQFSPTEFLAIONSCCTPPV	60			
1 MAHESESSRLGAVAGEPALGAGDASEDHPOYCAKCAQFIDPTEFLAIONKCTDPP	60			
61 MVIIGGENDPNSSASSAPRDEGHSRSQVMDTEHSPNDPSGSSGPPDPWGPERRGESS	120			
61 MVIIGGENDPNSSASSAPRDEGHSRSQVMDTEHSPNDPSGSSGPPDPWGPERRGESS	120			
121 GQFLVATGTAAGGGGGLTASPRTKIGAPLPDESPAPPPPPPPPPPGSGSHLNP	180			
121 GHLVATGTAAGGGGGLTASPRTKIGAPLPDESPAPPPPPPPPPPGSGSHLNP	180			
181 LEEELRVLDQROHOMQMTEDICROVLLGSLGQTAGAPASPELPTGTGAASKRPLPLF	240			
181 LEEELRVLDQROHOMQMTEDICROVLLGSLGQTAGAPASPELPTGTGAASKRPLPLF	240			
241 SPTRPAOTGTTA - SSSSSSSSSGAEPPKQAFHLVYHPLGSHPSVGVGGRSHKPPAP	299			
241 SPTRPAOTGTTA - SSSSSSSSSGAEPPKQAFHLVYHPLGSHPSVGVGGRSHKPPAP	299			
300 SPALPGSTDLIASPHLAFPTGTGTLAQCAGAAAGLEMAASPGLPKPKNGSELGYGEV	359			
301 SPALPGSTDLIASPHLAFPTGTGTLAQCAGAAAGLEMAASPGLPKPKNGSELGYGEV	359			
360 ISLELEPGGHHKCFCAKAVGSDSALOIHLSHTGERPKYCNVCGNRFTTGMKLVHFR	419			
361 MGLELPGGHHKCFCAKAVGSDSALOIHLSHTGERPKYCNVCGNRFTTGMKLVHFR	420			
420 HREKRYHVQNPVPVPEHLDPVITSSGLPYCGSVPEKAEEBAGTPGGGVERRPLVASTT	479			
421 HREKRYHVQNPVPVPEHLDPVITSSGLPYCGSVPEKAEEBAGTPGGGVERRPLVASTT	480			
480 ALSATESLTLTSTGTAVAAGLPTFNKFLVMKAVEPKSKADENTPPGSEGSIAIYADS	539			
481 ALSATESLTLTSTGTAVAAGLPTFNKFLVMKAVEPKSKADENTPPGSEGSIAIYADS	540			
540 GSATRQQLSLVSLPFWALITNLKSGSPFPVLPPLIGASPSSTSKLOOLVEKIDRQ	599			
541 SPATLMQSLKMLTSLPFWALITNLKSGSPFPVLPPLIGASPSSTSKLOOLVEKIDRQ	600			
600 GAVAAVASTASGAPTTSGAPAPSSASGSPNOCVICILRVLSCPALRLHYGHGGERPFCK	658			
601 GAVAAVASTASGAPTTSGAPAPSSASGSPNOCVICILRVLSCPALRLHYGHGGERPFCK	660			
659 VCGRASTGNTLAAHVGHTSPAPARONOSPICQKKTNTAVTLOOHVAMHIGQIPNG	718			
661 VCGRASTGNTLAAHVGHTSPAPARONOSPICQKKTNTAVTLOOHVAMHIGQIPNG	720			
719 SALSEGGAAOENSSEOSTASGPGSPFOPOSOQSPPEEEMS - EEEEDDEEEEDVDIDS	777			
721 TALPEGGAAOENSSEOSTASGPGSPFOPOSOQSPPEEELS - EEEEDDEEEEDVDIDS	780			
778 LAGRGSESGGEKALIVRGDSEEVYSGAEEFVATVLAAPTIVYEMDSNEKAPDHTLPPPPP	837			
781 LAGRGSESGGEKALIVRGDSEEVYSGAEEFVATVLAAPTIVYEMDSNEKAPDHTLPPPPP	840			

OY	838	PNDNDHQPMEQGTSDVSGAMEEBAKLEGISSPMALTOEGEGSTPLVEELNLPAMKK	897
Ddb	841	PDSDIDQPPMEDGSSGVLAGKEBGGCKPERSSSPASALTPEBEANISVTLYVELSLDEAMKK	900
OY	898	DPGESSGRKACEVCOSQSFPTQTALAEHQKTHPKRDPGLFTVCFCRGFLDRATLKRMHLA	957
Ddb	901	EPGSSSSRKACIEVCCOAFPSQALEEHOQKHPRKEGPLFTVCFCRGFLERATLKKMHLA	960
OY	958	HNOVPFPFPHGPQNATLSLVPGCSSSIPSGLSPPFRKDDPTMP	1002
Ddb	961	HHQVQPFAPHGPONIALSLVPGCSSPSTJSTGSLSPFRKDDPTIP	1005
<hr/>			
RESULT 4			
ID	AAE17968	standard; Protein; 1005 AA.	
XX	AAE17968;		
XX	07-MAY-2002	(first entry)	
XX	Human Sal2 protein mutant (G744R).		
XX	Human; proliferative disorder; tumor host range mutant virus; cancer;		
KW	T-HR mutant; Sal2 protein; ovarian tumour; mutant; mteIn.		
XX	Homo sapiens.		
OS	Synthetic.		
XX	Key	Location/Qualifiers	
FH	Misc-difference 744	/note= "Wild-type Gly substituted with Arg"	
FT			
FN	WC0200204596-AZ.		
PN	17-JAN-2002.		
PD			
XX	05-JUL-2001; 2001WO-US21354.		
PF			
XX	07-JUL-2000; 2000US-216723P.		
PR	19-MAR-2001; 2001US-081247I.		
PR	19-MAR-2001; 2001US-0812633.		
XX			
PA	(HARD) HARVARD COLLEGE.		
PA	(BGHM) BRIGHAM & WOMENS HOSPITAL INC.		
XX	Benjamin TL, LI D, Mok SC, Cramer DW, Ma Y;		
XX	WPI; 2002-164637/21.		
DR			
XX			
PT	Detecting protein involved in susceptibility to proliferative disease,		
PT	by infecting normal and abnormal proliferating cells with mutant virus,		
PT	detecting mutated protein allowing growth of mutant on abnormal cells		
PT	-		
XX			
PS	Example 7; Page -: 92pp; English.		
XX			
CC	The invention relates to a method for the identification of genes		
CC	and their encoded proteins involved in susceptibility to proliferative		
CC	disorders, including cancer using a tumor host range mutant (T-HR		
CC	mutant) virus. The invention also provides the use of sal2 genes and		
CC	proteins in methods of identifying a mammal having, or at a risk of		
CC	acquiring a proliferative disease. T-HR mutants are used to kill cancer		
CC	cells such as one carrying a Sal2 alteration. Transgenic and knockout		
CC	mouse comprising Sal2 nucleic acid are useful as research tools to		
CC	determine genetic and physiological features of cancer and for		
CC	identifying compounds that can affect ovarian and other tumours. The		
CC	present sequence is human Sal2 protein mutant (G744R).		
CC	Note: This sequence is not shown in the specification but is derived		
CC	from the human Sal2 wild-type protein shown as SEQ ID NO:1 (AAE17954)		
CC	in page 77-79 of the specification.		
XX			
SQ	Sequence 1005 AA:		

Query Match	88.1%	Score 4669.5	DB 23	Length 1005
Best Local Similarity	88.9%	Pred No. 7.1e-284		
Matches	893	Conservative	25	Mismatches 84; Indels 3; Gaps 3;
QY	1	MAOETGSSRLGPGCEPAERGGDASEEHHPOVCACCAOAFSDPTEFLAONSCSTDPPY	60	
DB	1	MAHESRSRLGPAEPALGGDASEEDHPQVCAKCAOFTDTEFLAONASTDDPY	60	
QY	61	MYITGGENPNSNSASAPRPEGHRSQVMDTHESNPDDSSCPDPPTWPERGEESS	120	
DB	61	MYITGGENPNSNSASASEPRPEGHNPQVMDTHESNPDDSSVPTDPTWPERGEESS	120	
QY	121	GQFLVATGTAAGGGGGLILASPRIGATPLPEESTPARPPPPPPPPGVSGHLNIPLI	180	
DB	121	GHLVATGTAAGGGGGLILASPRIGATPLPEESTPARPPPPPPPPGVSGHLNIPLI	180	
QY	181	LEELRLVLOQRHOMOMTEQICRQVLLLSGIVGVGAPASSELPGTAASSTRLLPLF	240	
DB	181	LEELRLVLOQRHOMOMTEQICRQVLLLSGIVGVGAPASSELPGTASTKLLPLF	240	
QY	241	SPFKPAPOTGKTTA-SSSSSSSSGAEPBPQAFPHLYHPLGSGQHPRSYGVGGRSHKPTPA	299	
DB	241	SPFKPAPOTKTTLASSSSSSSSSGAEPBPQAFPHLYHPLGSGQHPRSYGVGGRSHKPTPA	300	
QY	300	SPALPGSTDOLILASPHLAPGTTGLLAOCIGAARGLAEAAASPGILKPKNGSGELCYGV	359	
DB	301	SPALPGSTDOLILASPHLAFSTTGLLAOCIGAARGLAEATASPGILKPKNGSGELSYGV	360	
QY	360	ISLLEKGGHHKCFCKAKVSGDSALQHLKSHNGERYKCVKVCNRRFTRNKLKVHFR	419	
DB	361	MGPLEKGGHHKCFCKAKVSGDSALQHLKSHNGERYKCVKVCNRRFTRNKLKVHFR	420	
QY	420	HREKYPHVQNNPHRPHDLVYITSSGLPYGMSVPEKAEEDAGTGGVEKPLVASTT	479	
DB	421	HREKYPHVQNNPHRPHDLVYITSSGLPYGMSVPEKAEEDAAATPGGVEKPLVASTT	480	
QY	480	ALSTSELTLTSTSTRAVAGPLPTPKNFVLMKAEPKSKADENTPGSEGSINGVAAS	539	
DB	481	ALSTSELTLTSTSTAGATATAGPLPFPNFVLMKAEPKKADENTPGSEGSINGVAAS	540	
QY	540	GSATRMQLSVTLSPSWALLTNHLKTSGFPPFVLPPLGASPSBSTKLQOLVEKIDRQ	599	
DB	541	STALMLQSLKMTLPSWALLTNHFKSTGSPRLPCARALGASPSBSTKLQOLVEKIDRQ	600	
QY	600	GAVAVASTAGAPTTAPAPSSA-SGRNOCYICRLVSCPALRLHYHQHGERFPCK	658	
DB	601	GAVAVTSAAGAPTTAPAPSSASAGNOCYICRLVSCPALRLHYHQHGERFPCK	660	
QY	659	VCGAFTSRGULRAHFVGHKTPARAONSCITCKKFTNATTLQOHVNMHLGQILPNG	718	
DB	661	VCGAFTSRGULRAHFVGHKSPARAONSCITCKKFTNATTLQOHVNMHLGQILPNG	720	
QY	719	SALSEGGGAOENSSSEOSTASGPGSFPOPOSOQSPSEEMS-EDEEDEDEEDVDDED	777	
DB	721	TALPEGGGAOENSGESTVSGARFPQOOSQSPSEEBELSEEBEDEDEEDVDDED	780	
QY	778	LAGGSEBEGEKAISVYRDSSEFVGAEEFVAATVTKEMDSNEKAPQHLTPPPPP	837	
DB	781	LAGGSEBEGEKAISVYRDSSEFASAAEEFVVAALATGKEMDSNEKTIQOSSLPPPP	840	
QY	838	PDNLDPHPMEOGTSDVGVGAMEEBKLDEISSPMALTOEGECTSTPLVELNLPEAMK	897	
DB	841	PDNLDPHPMEOGSSGVLAGKEGDKPEKSSSPASALTPREGAITSVTLVELLLEAMK	900	
QY	898	DPGESSGRKACEVCGSEPTOTALBEOHKTNRKDGFLPTCYFCRQGLDRAATLKKHMLA	957	
DB	901	EPGESSSRKACEVCGAEPFQALBEOHKTNRHEGFLPTCYFCRQGLLEATATLKKHMLA	960	
QY	958	HHOVPFAPHGPONTATLTVPGCSSJPSGLSPRRKDDPMR 1002		
DB	961	HHOVPFAPHGPONTALIASLVPGCSSJTSGLSPRRKDDPMR 1005		

RESULT 5
AA079822
ID AAM79822 standard; Protein; 1019 AA.
XX
AC AAM79822;
XX
XX 06-NOV-2001 (first entry)
XX
XX Human protein SEQ ID NO 3468.
XX
XX
XX Human: cytokine; cell proliferation; cell differentiation; gene therapy;
XX vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
XX tissue growth factor; immunomodulatory; cancer; leukaemia;
XX nervous system disorder; arthritis; inflammation.
XX
XX Homo sapiens.
XX
XX W0200157190-R2.
XX
XX
XX 09-AUG-2001.
XX
XX 05-FEB-2001; 2001WO-US04098.
XX
XX 03-FEB-2000; 2000US-0496914.
XX 27-APR-2000; 2000US-0560875.
XX 20-JUN-2000; 2000US-0598075.
XX 19-JUL-2000; 2000US-0620325.
XX 01-SEP-2000; 2000US-0654936.
XX 15-SEP-2000; 2000US-0663561.
XX 20-OCT-2000; 2000US-0693325.
XX 30-NOV-2000; 2000US-0728422.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Tang YT, Liu C, Dirmacac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;
XX Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
XX Xue AJ, Yang Y, Wejhrman T, Goodrich R;
XX
XX WPI: 2001-4/6283/51.
XX N-PSDB: AAK52955.
XX
XX Nucleic acids encoding polypeptides with cytokine-like activities,
XX useful in diagnosis and gene therapy -
XX
XX Claim 20; Page 347-348; 6221PP; English.
XX
XX The invention relates to polynucleotides (AAK51456-AAK53435) and the
XX encoded polypeptides (AAM78323-AA080302) that exhibit activity elating to
XX cytokine, cell proliferation or cell differentiation or which may induce
XX production of other cytokines in other cell populations. The
XX polynucleotides and polypeptides are useful in gene therapy, vaccines or
XX peptide therapy. The polypeptides have various cytokine-like activities,
XX e.g. stem cell growth factor activity, haematopoiesis regulating
XX activity, tissue growth factor activity, immunomodulatory activity and
XX activity/inhibin activity and may be useful in the diagnosis and/or
XX treatment of cancer, leukaemia, nervous system disorders, arthritis and
XX inflammation.
XX NOTE: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
XX (AAM00020) are omitted as the relevant pages from the sequence listing
XX were missing at the time of publication.
XX
XX Sequence 1019 AA:

Query Match 87.3%; Score 4625.5; DB 22; Length 1019;
Best local similarity 89.2%; Pred. No. 4.1e-281;
Matches 883; Conservative 23; Mismatches 81; Indels 3; Gaps 3;
Y 16 GEPAREGDSAEENHPQCAKCCAFSDPTFEFLAHONSCSTDPPVWYIIGQENPNSNSA 75
Y 30 GPSASENGDSAEENHPQCAKCCAFSDPTFEFLAHONACSTDPPVWYIIGQENPNSNSA 89
Y 76 SSAPRREGHSRQVMDTSHSNRPDSSGSRPPPTWGPERRGESSSGQFLVAATGTAAAGG 135

Db 90 SSAPRREGHNHPQVMDTSHSNRPDSSGSRPPPTWGPERRGESSSGHFLVAATGTAAAGG 149
QY 136 GGLILASPKIGATPLPEESTPAPPPPPPPPPGVSGHLNPLILEELRVLQORIQM 195
Db 150 GGLILASPKIGATPLPEESTPAPPPPPPPPPGVSGHLNPLILEELRVLQORIQM 209
QY 196 QMTEQICROYLLIGSGQVGAAPASPSSELPGTGAASSYRKPLPLPSPIKPAQIGKTTA-S 254
Db 210 QMTEQICROYLLIGSGQVGAAPASPSSELPGTGAASSYRKPLPLPSPIKPAQIGKTTA-S 269
QY 255 SSSSSSSSGAEPPKQAFHLYHPILGSOHPFESVGVGRSHKPPAPSPALPGSTQDLIAP 314
Db 270 SSSSSSSSGAEPPKQAFHLYHPILGSOHPFESVGVGRSHKPPAPSPALPGSTQDLIAP 329
QY 315 HLAEPGTTGLLAOCIGAAAGLEAASBGLKPKNGSELGVEYISLEKPGRRKCRF 374
Db 330 HLAEPGTTGLLAOCIGAAAGLEAASBGLKPKNGSELGVEYISLEKPGRRKCRF 389
QY 375 CAKVFSGDSALQIHLSSHTEGEPYKCNVCGNRPTRGNLKYHFRHREKYPHYQNMHPY 434
Db 390 CAKVFSGDSALQIHLSSHTEGEPYKCNVCGNRPTRGNLKYHFRHREKYPHYQNMHPY 449
QY 435 PEHLDVITSSGLPYGMSVPEKAEEDAGTPGGVERKPLVAATLSTESLTLTGT 494
Db 450 PEHLDVITSSGLPYGMSVPEKAEEDAGTPGGVERKPLVAATLSTESLTLTGT 509
QY 495 STAVAPGLPTFNKFLMKAVERPASKADENPPGSEGSIAIYVADSGSATRMOLSKLYTSL 554
Db 510 STAVAPGLPTFNKFLMKAVERPASKADENPPGSEGSIAIYVADSGSATRMOLSKLYTSL 569
QY 555 PSMALLTNHLKSTGSPFPVLEPIGASPSETSKLOQLVEKIDRGAVAVAASGAPTT 614
Db 570 PSMALLTNHLKSTGSPFPVLEPIGASPSETSKLOQLVEKIDRGAVAVAASGAPTT 629
QY 615 SAPAPSSSA-SCPNOCVILRVLSCPRALRLHGOHGERPFCCKVCGRAFTSGNLRH 673
Db 630 SAPAPSSSA-SCPNOCVILRVLSCPRALRLHGOHGERPFCCKVCGRAFTSGNLRH 689
QY 674 FVGHKTPPARAONSCPIQOKFTNAVTLOOHVHMHLGGQIPNGSALSEGGAQOENS 733
Db 690 FVGHKTPPARAONSCPIQOKFTNAVTLOOHVHMHLGGQIPNGSALSEGGAQOENS 749
QY 734 EQSTASGSPFPQOSQOQSPSEEM-EEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDE 792
Db 750 EQSTASGSPFPQOSQOQSPSEEM-EEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDE 809
QY 793 VRGDEEVSAGAEVATVAAPTTVKEMDSNEKAPQHTLPPPPPPDNLDPQMEQCTS 852
Db 810 VRGDEEVSAGAEVATVAAPTTVKEMDSNEKAPQHTLPPPPPPDNLDPQMEQCTS 869
QY 853 DVSGAMEEAKLEGISPPAALTOEGEGSTPLVEELNPPAMKDDPSSSRKACEYCG 912
Db 870 GVLGKEGEGKREKRSSPASALTPCEGATVTVLEELSLOEAMKREGESSSRKACEYCG 929
QY 913 QSPFQALAEHQKTHPKDQPLTCVFCROGLDRATLKKHMLLAHNOVPFAPHPONI 972
Db 930 QAFPSOALAEHQKTHPKDQPLTCVFCROGLDRATLKKHMLLAHNOVPFAPHPONI 989
QY 973 ATLSTVPGCSSIPSPGSLPPPKDDPTM 1002
Db 990 AALSTVPGCSSIPSPGSLPPPKDDPTM 1019

RESULT 6
AAM78838
ID AAM78838 standard; Protein; 1007 AA.
XX
XX AAM78838;
XX
XX 06-NOV-2001 (first entry)
XX
XX Human protein SEQ ID NO 1500.

Human: cytokine; cell proliferation; cell differentiation; gene therapy;
vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
tissue growth factor; immunomodulatory; cancer; leukaemia;
nervous system disorder; arthritis; inflammation.
Homo sapiens.
MO200157190-A2.
09-AUG-2001.
05-FEB-2001; 2001WO-US04098.
03-FEB-2000; 2000US-046914.
27-APR-2000; 2000US-0560875.
20-JUN-2000; 2000US-0598075.
19-JUL-2000; 2000US-0620325.
01-SEP-2000; 2000US-0654936.
15-SEP-2000; 2000US-0663561.
20-OCT-2000; 2000US-0693325.
30-NOV-2000; 2000US-0728422.
(HYSEQ-) HYSEQ INC.
Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;
Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
Xue AU, Yang Y, Wejhtman T, Goodrich R;
WPI: 2001-476283/51.
N-PSDB; AAK51971.
Nucleic acids encoding polypeptides with cytokine-like activities,
useful in diagnosis and gene therapy -
Claim 20; Page 3790-3792; 6221pp; English.
The invention relates to polynucleotides (AAK51456-AAK53435) and the
encoded polypeptides (AAK78323-AAK80302) that exhibit activity elating to
cytokine, cell proliferation or cell differentiation or which may induce
production of other cytokines in other cell populations. The
polynucleotides and polypeptides are useful in gene therapy, vaccines or
peptide therapy. The polypeptides have various cytokine-like activities,
e.g. stem cell growth factor activity, haematopoiesis regulating
activity, tissue growth factor activity, immunomodulatory activity and
activity/inhibit activity and may be useful in the diagnosis and/or
treatment of cancer, leukaemia, nervous system disorders, arthritis and
inflammation.
Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
(AAK80020) are omitted as the relevant pages from the sequence listing
were missing at the time of publication.
Sequence 1007 AA:
Query Match 87.1%; Score 4615.5; DB 22; Length 1007;
Best Local Similarity 89.0%; Pred. No. 1.7e-280;
Matches 881; Conservative 24; Mismatches 82; Indels 3; Gaps 3;
16 GEPAEKGGDAEEHHPVCAKCAQFSPTERLAHQNSCTDPVAVYITGGQENPSSSA 75
18 GEPASERGGDAEEHHPVCAKCAQFPTERLAHQNSCTDPVAVYITGGQENPSSSA 77
76 SSAPREHGHSRQVMDREHNSPPDSGSGPPPTWGPERRGEESSGOFLVAATGTAAGG 135
78 SSAPREHGHSRQVMDREHNSPPDSGSGVPPTWGPERRGEESSGFLVAATGTAAGG 137
136 GGLILASPKLGATPLPESTAPAPPPPPPPGVSGHLNIPLLLEELRVLOQRQHOM 195
138 GGLILASPKLGATPLPESTAPAPPPPPPPGVSGHLNIPLLLEELRVLOQRQHOM 197
196 QMTEQICROYLLLSLQGTVCAPASPELPGTGAASSTKPLPLFSFIKPAQTKTKTA-S 254
198 QMTEQICROYLLLSLQGTVCAPASPELPGTGAASSTKPLPLFSFIKPAQTKTKTASS 257

255 SSSSSSSGAEPPKQAFHLHYHDLGSGQHPPSVGVGSRSHKPTPAPSPALPGSTDLIASP 314
258 SSSSSSSGAEPPKQAFHLHYHDLGSGQHPPSVGVGSRSHKPTPAPSPALPGSTDLIASP 317
315 HLAFFPTGTLAAQCGAARGLEATAAPSPGLLPRKNGSGELGVEYISSLEKKGRRKCRF 374
318 HLAFFPTGTLAAQCGAARGLEATAAPSPGLLPRKNGSGELGVEYISSLEKKGRRKCRF 377
375 CAKVFGSDALQIHLSHNGERYKCNVCGNRFTTGRNLKVFHRRHREKYPHYOMNPHYV 434
378 CAKVFGSDALQIHLSHNGERYKCNVCGNRFTTGRNLKVFHRRHREKYPHYOMNPHYV 437
435 PEHLDYVITSSGLPYGMSVPEPKAEEDAGTPGGVERKPLVASTALATESITLLSTGT 494
438 PEHLDYVITSSGLPYGMSVPEPKAEEDAGTPGGVERKPLVASTALATESITLLSTGT 497
495 STAVAPGLPTFNKFLVLMKAVEPKSKADENTPPGSEGSALGAVADGSATRMQLSKVLSL 554
498 GTATAPGLPAFNKFLVLMKAVEPKSKADENTPPGSEGSALGAVADGSATRMQLSKVLSL 557
555 PSMALLTNHLKSTGSPFPVLEPLGASPSFTSKLOLVEKIDROGAVAVASTAGAPT 614
558 PSMALLTNHLKSTGSPFPVLEPLGASPSFTSKLOLVEKIDROGAVAVASTAGAPT 617
615 SAPAPSSSA-SGPNQVICTRLVLSCPRALRLHYGONGGERPEKCKVCGRAFSTRNLRAH 673
618 SAPAPSSSASGPNQVICTRLVLSCPRALRLHYGONGGERPEKCKVCGRAFSTRNLRAH 677
674 FVGHKTSPPARAONSCPTCOCKKFTNAVTLOQHVRNHLGQITNGSGALSEGGAAQENSS 733
678 FVGHKTSPPARAONSCPTCOCKKFTNAVTLOQHVRNHLGQITNGSGALSEGGAAQENSS 737
734 EOSTAGSPSPPOQSOOPSPSEEMS--EEDDEDEEDVDTEDESLAGRSESGEKAIS 792
738 EOSTAGSPSPPOQSOOPSPSEEMS--EEDDEDEEDVDTEDESLAGRSESGEKAIS 797
793 VRGDSSEVSAGAEENVATVAPRTVYKEMDSNEKADONLTPPPPPDNIDHPQMEQGT 852
798 VRGDSSEVSAGAEENVATVAPRTVYKEMDSNEKADONLTPPPPPDNIDHPQMEQGT 857
853 DVSAGMEERKLEGISPMALTOEGESTSTPLVEELNPEAMKKGPGSSGRKACVCG 912
858 GYVGKREGEKPERKSSPASLTPREGATSVLVEELSLQEAARKKEGESSSRKACEVCG 917
913 QSPPTOTALFEHOKTHPRDGLFTCVFCROGFLDRATLKKHMLAHNOVPPFAPHGONI 972
918 QAPPSQALAEHOKTHPRDGLFTCVFCROGFLDRATLKKHMLAHNOVPPFAPHGONI 977
973 ATLSLVPGCSSIPSPGLSPPPRKDDPTMP 1002
978 ATLSLVPGCSSIPSPGLSPPPRKDDPTMP 1007
RESULT 7
AAB93193
ID AAB93193 standard; Protein; 813 AA.
XX AAB93193:
AC 26-JUN-2001 (first entry)
XX Human protein sequence SEQ ID NO:12150.
XX Human: primer; detection; diagnosis; antisense therapy; gene therapy.
OS Homo sapiens.
XX EP1074617-A2.
XX EP1074617-A2.
XX 07-FEB-2001.
XX 28-JUL-2000; 2000EP-0116126.
PF

AC	AAM56858,	
XX		
DT	05-NOV-2001	(first entry)
DE	Human brain expressed single exon probe encoded protein SEQ ID NO: 28963.	
XX		
XX	Human: brain expressed exon; gene expression analysis; probe:	
KW	microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;	
KW	epilepsy; cancer.	
XX		
OS	Homo sapiens.	
PN	MO200157275-A2.	
PD		
PE	09-AUG-2001.	
PR	30-JAN-2001; 2001WO-US00667.	
XX		
PR	04-FEB-2000; 2000US-0180312.	
PR	26-MAY-2000; 2000US-0207456.	
PR	30-JUN-2000; 2000US-0608408.	
PR	03-AUG-2000; 2000US-0632366.	
PR	21-SEP-2000; 2000US-0234687.	
PR	27-SEP-2000; 2000US-0236359.	
XX	04-OCT-2000; 2000GB-0024263.	
PA	(MOL-) MOLECULAR DYNAMICS INC.	
PI	Penn SG, Hanzel DK, Chen W, Rank DR;	
DR	WPI: 2001-483446/52.	
XX		
PT	Single exon nucleic acid probes for analyzing gene expression in human	
PT	brains -	
XX		
PS	Example 4: SEQ ID NO: 28963; 650bp + Sequence Listing; English.	
XX		
CC	The present invention provides a number of single exon nucleic acid	
CC	probes which are derived from genomic sequences expressed in the human	
CC	brain. They can be used to measure gene expression in brain cell samples,	
CC	which may enable the diagnosis and improved treatment of nervous system	
CC	diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,	
CC	epilepsy and cancers. The present sequence is a protein encoded by one of	
CC	the probes of the invention.	
XX		
SQ	Sequence 330 AA:	
	Query Match 9.1%; Score 482.5; DB 22; Length 330;	
	Best Local Similarity 35.7%; Pred. No. 2,5e+22;	
	Matches 131; Conservative 37; Mismatches 96; Indels 101; Gaps 13.	
OY	531 SAAGVAGSGSATRMOLSKLTSLPMSALLTNHLKSTGSPFPFVLEPLDASPSETSKIQ 590	
Db	I I : : : : I I : : : : I I : : : : I I I I I I	
	4 SPADCGPAGATFT-----TNPDLPLMSEQPKA--KFFPGGLD--SAOASETSKIQ 52	
OY	591 OLVEKIDRCGAANAFASTASCAPPTSARPARSSASGNCOVICLRVLSCPALRLHYGGH 650	
Db	I I I I I I : : : : I I : : : : I I : : : : I I I I I I	
	53 OLVENIDDK-----ATDPNECITCHRVSQSALKMHYRTHT 89	
OY	651 GERPFCKVCYGRAFTGRNLRHAHEYGHKTSPPARAQNSCPICKKFNNATYLQQHVRAHL 710	
Db	I I I I I I : : : : I I I I I I : : : : I I : : : : I I I I I I	
	90 GERPFCKVICGRAFTTGKLNKLTHYSVHRAMPRLRVQHSCPICQKKFTNAVYLLQDHIRHM 149	
OY	711 GGQIIPN-----GGSALISEGGAAQEN-----SSEOSTASGPSPFO-PQSOQPSPRE 755	
Db	I I I I I I : : : : I I : : : : I I : : : : I I I I I I	
	150 GGQIIPNTPVDSDYSESMESTDYGFDEKNFDLDNFSDENNEDPREGSIIPRKSADAQOD 209	
OY	756 E-----EMSE-----EEEDDEEEDVTDDSDLAGRG-- 782	
Db	II I I I I : : : : I I : : : : I I I I I I : : : :	
	210 SLSSSPFLPLEKSSIALLENQKMIMAGLAQLQASLKYVNGSTEGGVLTNDSSVGSDM 269	
OY	783 -SSSGGEKAISVRGDSEEV---SGAEDEVATSVAADPTVTKEMDSNEKAPQHTLPDP--- 834	
	I I I I I I I I : : : : I I : : : : I I I I I I : : I	

[illegible]

[illegible]

QY	531	SALATGAVDSSGATITMQLSKLVTSLFSPMALLTNHLKSTGSPFFVYVLEPLGASPETSKLQ	590
Db			
	4	SPADDCGPGAGATATTF-----TNPLLPLMSQEFKA--KFPFGGLD--SQAQSETSKLQ	52
QY	591	QLVEKIDROGAVVAASTASGAPPTTSAPAPSSASGPNOCVIGLCRLVSCPPALRLHYGQHG	650
Db			
	53	QIVENIDKK-----ATDPNECTICHRVLSQSLAKMHTRTHT	89
QY	651	GERPFCKVCGRATSTRGNLRAHFVGHKTSAPARAONSPICQKKTNAVTLQGHVHML	710
Db			
	90	GERPFCKICGRATFTKNTLKHYSVHRAMPRLRVQHSPCICQKKTNAVTLQGHVHML	149
QY	711	GGGIQPN-----GGSALSEGGAQEN-----SSDSTAGSPSPQ--POSQSPSE	755
Db			
	150	GGGIQPNTPVPDSTYSSESMSDTSFDEKFNDDIDNFSDENMEDCPESGISPTPKSADASQD	209
QY	756	E-----EMSE-----EEEEDEEEEDVTDEDLSAGRG--	782
Db			
	210	SLSSPPLPLEMSSIIVALENQMKMINAGLAEQLQASLKSVENGSIEDDYLINDSSVGGDM	269
QY	783	-SSSGCKALSVKCDSEV---SCAEEEVANISVAAPTIVKEMDSNEKAPQHTLPP---	834
Db			
	270	ESGSAGSPALISESTSSQALSPSNSTQEFHKS-----PSIEEKQRAVPSSEFANG	319
QY	835	-PPPPDN 840	
Db			
	320	LSPTPVN 326	
RESULT 17			
AA	29568		
ID	AA029568	standard; Protein; 330 AA.	
XX	AA		
AC	AA029568;		
XX			
DT	17-OCT-2001	(first entry)	
XX			
DE	Peptide #3605 encoded by probe for measuring placental gene expression.		
XX			
KW	Probe; microarray; human; placenta; antenatal diagnosis;		
XX			
OS	Homo sapiens.		
PN	WO200157272-A2.		
XX			
PD	09-AUG-2001.		
XX			
PE	30-JAN-2001; 2001WO-US00663.		
XX			
PR	04-FEB-2000; 2000US-0180312.		
PR	26-MAY-2000; 2000US-0207456.		
PR	30-JUN-2000; 2000US-0606408.		
PR	03-AUG-2000; 2000US-0632366.		
PR	21-SEP-2000; 2000US-0234687.		
PR	27-SEP-2000; 2000US-0236359.		
PR	04-OCT-2000; 2000GB-0024263.		
XX			
PA	(MOLE-) MOLECULAR DYNAMICS INC.		
XX			
PI	Penn SG, Hanzel DK, Chen W, Rank DR;		
XX			
DR	WPI; 2001-488897/53.		
XX			
PT	Human genome-derived single exon nucleic acid probes useful for		
PT	analyzing gene expression in human placenta -		
XX			
PS	Claim 27; SEQ ID No 29837; 654bp; English.		
XX			
CC	The present invention relates to single exon nucleic acid probes (SNP: see AI13115-AI157546). The present sequence is a microarray encoded by one such probe. The probes are useful for producing a microarray for		

hyaline membrane disease.

Homo sapiens.

WC200186003-A2.

15-NOV-2001.

30-JAN-2001; 2001WO-0500665.

04-FEB-2000; 2000US-180312P.
26-MAY-2000; 2000US-207456P.
30-JUN-2000; 2000US-0608408.
03-AUG-2000; 2000US-0632366.
21-SEP-2000; 2000US-234687P.
27-SEP-2000; 2000US-236359P.
04-OCT-2000; 2000GB-0024263.

(MOLE-) MOLECULAR DYNAMICS INC.

Penn SG, Hanzel DK, Chen W, Rank DR;
WPI; 2002-114183/15.

Spatially-addressable set of single exon nucleic acid probes, used to measure gene expression in human lung samples -

Claim 27; SEQ ID NO 28520; 634pp; English.

The invention relates to a spatially-addressable set of single exon nucleic acid probes for measuring gene expression in a sample derived from human lung comprising single exon nucleic acid probes having one of 12614 nucleic acid sequences mentioned in the specification, or their complements or the 12387 open reading frames derived from the 12614 probes. Also included are a microarray comprising the novel set of probes; the novel set of probes which hybridise at high stringency to a nucleic acid expressed in the human lung; measuring gene expression in a sample derived from human lung, comprising (a) contacting the array with a collection of detectably labeled nucleic acids derived from human lung mRNA, and (b) measuring the label detectably bound to each probe of the array; identifying exons in a eukaryotic genome, comprising (a) algorithmically predicting at least one exon from genomic sequences of the eukaryote; and (b) detecting specific hybridisation of detectably labeled nucleic acids from eukaryote lung mRNA, to a single exon probe, having a fragment identical to the predicted exon, the probe is included in the above mentioned microarray; assigning exons to a single gene, comprising (a) identifying exons from genomic sequence by the method above and (b) measuring the expression of each of the exons in several tissues and/or cell types using hybridisation to a single exon microarrays having a probe with the exon, where a common pattern of expression of the exons in the tissues and/or cell types indicates that the exons should be assigned to a single gene; a peptide comprising one of 12011 sequences, mentioned in the specification, or encoded by the probes/open reading frames (ORF). The probes are used for gene expression analysis, and for identifying exons in a gene, particularly using human lung derived mRNA and for the study of lung diseases such as asthma, lung cancer, chronic obstructive pulmonary disease (COPD), interstitial lung disease (ILD), familial idiopathic pulmonary fibrosis, neurofibromatosis, tuberous sclerosis, Gaucher's disease, Niemann-Pick disease, Hermansky-Pudlak syndrome, sarcoidosis, pulmonary haemangioendothelioma, lymphangioleiomyomatosis, pulmonary alveolar proteinosis, Karsenger syndrome, fibrocystic pulmonary dysplasia, primary ciliary dyskinesia, pulmonary hypertension and hyaline membrane disease. The present sequence is a peptide/protein encoded by a single exon probe of the invention.

Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at [ftp.wipo.int/pub/published_pct_sequences](http://wipo.int/pub/published_pct_sequences).

Sequence 330 AA:

9.1%: Score:82.5; DB 23: Length 330:

Best Local Similarity	35.7%;	Pred. No. 2,5e-22;	
Matches 131;	Conservative 37;	Mismatches 98;	Indels 101; Gaps 13;
OY	531	SATAGVADSGSARFMOLSKLVTSLPSMALLTNHLKSTGSPFPPEVLEPLGASPSSTKLQ	590
		: : : : : : : : : : : : : : : : : : : :	
Db	4	SPRADCGPACASATTF-----TNPLPLMSEQFKA--KEPFGILD--SAQASSTKLQ	52
OY	591	QLVEKIDRCQAVAVASTASCAPTTSAPAPSSASGPNOCVICLRVLSCPALRIHYGCHG	650
		: : : : : : : : : : : : : : : : : : : :	
Db	53	QVENVNDKK-----ATDPNECILLGRVYSCOSALKMRYRTH	89
OY	651	GERPEKCKYCGRAFTSGNLRHAFVGHKTSAPARARNSCPICOKKFTNAYTLIOHVRMHL	710
		: : : : : : : : : : : : : : : : : : : :	
Db	90	GEPEPKCKICIGRAFTTGKLTHTSVHAMPRLRVHOSCPICKKFTNAYVLOQHIRMH	149
OY	711	GGQIPN-----GGSALSEGGAQOEN-----SSEOSTASGPPSPFO--POSQDPSPE	755
		: : : : : : : : : : : : : : : : : : : :	
Db	150	GGQIPNTPVPDYSSESMSHSDGSPDEKTFDDLDNNSDEMMEDECPESGIPPTPKSADASQD	209
OY	756	E-----EMSE-----EEDDEEEDDYDEDSLARG--	782
		: : : : : : : : : : : : : : : : : : : :	
Db	210	SLSSSEPLFLEMSIAALENQMKNINAGLAEOQLASLKSVENGSIEGDVLTINDSSVGGDM	269
OY	783	-SESGGEKATSVKQDSPEV---SGAEEVATSVAAPTVKEMDSNKAPOHTLPP----	834
		: : : : : : : : : : : : : : : : : : : :	
Db	270	EQQSASCPAISESTSSQALSPSNTQEFHKS-----PSIEKPKQAVAPSEFPANG	319
OY	835	-PPPPDN 840	
		: : : : :	
Db	320	LSPTPYN 326	
RESULT 20			
ABB32272			
XX ID	ABB32272	standard; peptide; 336 AA.	
XX AC	ABB32272;		
XX DT	01-FEB-2002	(first entry)	
XX DE	Peptide #4923	encoded by breast cell single exon nucleic acid probe.	
XX KW	Human; microarray; single exon probe; gene expression; breast; disease; cancer.		
XX OS	Homo sapiens.		
XX WO	WO200157271-A2.		
XX PD	09-AUG-2001.		
XX PF	30-JAN-2001; 2001MO-US00662.		
XX PR	04-FEB-2000; 2000US-0180312.		
XX PR	26-MAY-2000; 2000US-0207456.		
XX PR	30-JUN-2000; 2000US-0608408.		
XX PR	03-AUG-2000; 2000US-0632366.		
XX PR	21-SEP-2000; 2000US-0234687.		
XX PR	27-SEP-2000; 2000US-0236359.		
XX PR	04-OCT-2000; 2000GB-0024263.		
XX PA	(MOLE-) MOLECULAR DYNAMICS INC.		
XX PI	Penn SG, Hanzel DK, Chen W, Rank DR;		
XX DR	WPI; 2001-496933/54.		
XX PT	New spatially-addressable set of single exon nucleic acid probes,		
XX PT	useful for measuring gene expression in sample derived from human		
XX PT	breast, comprises number of single exon nucleic acid probes -		
XX	Claim 27; SEQ ID NO 15240; 327bp + sequence listing; English.		

CC The invention relates to a spatially-addressable set of single exon
CC nucleic acid probes for measuring gene expression in a sample derived
CC from human breast and BT 474 cells. The method involves contacting
CC the probes with a collection of detectably labelled nucleic acids
CC derived from mRNA of human breast, and then measuring the label
CC bound to each probe of the microarray. The probes are useful for
CC verifying the expression of regions of genomic DNA predicted to
CC encode proteins. They are useful for gene discovery, and for
CC determining predisposition and/or prognosing breast disease. Gene
CC expression analysis is useful for assessing the toxicity of chemical
CC agents on cells. The microarray of this invention presents a far greater
CC diversity of probes for measuring gene expression, with far less bias
CC than expressed sequence tag microarrays. The method is suitable for
CC rapid production of functional information from genomic sequence. The
CC present sequence is a peptide encoded by a single exon nucleic acid
CC probe of the invention.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pcl_sequences.

XX Sequence 336 AA:

Query Match 8.3%; Score 438.5; DB 22; Length 336;
Best Local Similarity 30.4%; Pred. No. 1.5e-19;
Matches 128; Conservative 42; Mismatches 128; Indels 123; Gaps 13;

XX 510 LMKAVEPKSKADENTP-----PSEGS-ATAGVADSGSATRMQLSKLVTSLPSWALL 560

DB 5 LSSGTPKDLTGSLPGDLQPPSPSESGPTRLPGVGPVNSPR----- 48

XX 561 TNLKSTGSPFPYVLEPLGASPSFETSKLOOLVEKIDRQGANAVASTAGAPTTAPAPS 620

DB 49 -----AGGFQSGTPEP-----GSETLKLQQLVENDIK----- 76

XX 621 SSASGPNOCVICTLRLVLSCPRALRLHYGOHGERPFCKVCYGRAFSRGNLRAHFVGHKTS 680

DB 77 -ATDPNECLICHVLSGQSSLKMHYRTHTGERPFQCKICGRAFSRGNLKTHLGVHRTN 135

XX 681 PAARQNSCPICQKKTNAVTLQOHVNRHMLGGOITPNGSALSSEGGAAOENSSQOSTASG 740

DB 136 TSIKTOHSCPTICQKKTNAVTLQOHVNRHMLGGOITP----- 171

XX 741 PGSPFOPOQSPPEEEMSEEEDEEDVDTEDESLAGRSESGEKAI-----SV 793

DB 172 -----TLPENPC-----DFTGSEPMT--VGENGSTGALCHDVIEST 207

XX 794 RGDSEEVSGAEEEVATVAAPTIVKEMDSNEK---APQHTLPP--PPPDNLDPOR 846

DB 208 --DVEEVS--SQEAPSSSSKVPTRPLPSIHASPTLGFAMMASLDAPGKVPAPFNLOQGS 264

XX 847 MEOGTSDVSGAMEEBAKLEGISPMALTOEGEGTSTPLVEELNLPAMKKDPRESSGRK 906

DB 265 RENGSVESDGLTNDSSSLMGDOEYQSRSPDILETTTSFOALSPANQAESTIKSPDAGSK 324

XX 907 A 907

DB 325 A 325

XX RESULT 21

XX ABB37530 ID ABB37530 standard; Peptide; 336 AA.

XX AC ABB37530;

XX DT 04-FEB-2002 (first entry)

XX DE Peptide #5036 encoded by human foetal liver single exon probe.

XX KW Human; foetal liver; gene expression; single exon nucleic acid probe.

XX OS Homo sapiens.

PN W0200157277-A2.

XX 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US00669.

XX 04-FEB-2000; 2000US-0180312.

XX 26-MAY-2000; 2000US-0207456.

XX 30-JUN-2000; 2000US-0608408.

XX 03-AUG-2000; 2000US-0632366.

XX 21-SEP-2000; 2000US-0234687.

XX 27-SEP-2000; 2000US-0236359.

XX 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-483447/52.

XX Human genome-derived single exon nucleic acid probes useful for

XX analyzing gene expression in human fetal liver -

XX Claim 27; SEQ ID NO 30165; 639pp + sequence listing; English.

XX The invention relates to a single exon nucleic acid probe for

XX measuring human gene expression in a sample derived from human foetal

XX liver. The single exon nucleic acid probes may be used for predicting,

XX measuring and displaying gene expression in samples derived from human

XX fetal liver. The present sequence is a peptide encoded by a single exon

XX nucleic acid probe of the invention.

XX Note: The sequence data for this patent did not form part of the

XX printed specification, but was obtained in electronic format directly

XX from WIPO at ftp.wipo.int/pub/published_pcl_sequences.

XX Sequence 336 AA:

XX Query Match 8.3%; Score 438.5; DB 22; Length 336;

XX Best Local Similarity 30.4%; Pred. No. 1.5e-19;

XX Matches 128; Conservative 42; Mismatches 128; Indels 123; Gaps 13;

XX 510 LMKAVEPKSKADENTP-----PSEGS-ATAGVADSGSATRMQLSKLVTSLPSWALL 560

DB 5 LSSGTPKDLTGSLPGDLQPPSPSESGPTRLPGVGPVNSPR----- 48

XX 561 TNLKSTGSPFPYVLEPLGASPSFETSKLOOLVEKIDRQGANAVASTAGAPTTAPAPS 620

DB 49 -----AGGFQSGTPEP-----GSETLKLQQLVENDIK----- 76

XX 621 SSASGPNOCVICTLRLVLSCPRALRLHYGOHGERPFCKVCYGRAFSRGNLRAHFVGHKTS 680

DB 77 -ATDPNECLICHVLSGQSSLKMHYRTHTGERPFQCKICGRAFSRGNLKTHLGVHRTN 135

XX 681 PAARQNSCPICQKKTNAVTLQOHVNRHMLGGOITPNGSALSSEGGAAOENSSQOSTASG 740

DB 136 TSIKTOHSCPTICQKKTNAVTLQOHVNRHMLGGOITP----- 171

XX 741 PGSPFOPOQSPPEEEMSEEEDEEDVDTEDESLAGRSESGEKAI-----SV 793

DB 172 -----TLPENPC-----DFTGSEPMT--VGENGSTGALCHDVIEST 207

XX 794 RGDSEEVSGAEEEVATVAAPTIVKEMDSNEK---APQHTLPP--PPPDNLDPOR 846

DB 208 --DVEEVS--SQEAPSSSSKVPTRPLPSIHASPTLGFAMMASLDAPGKVPAPFNLOQGS 264

XX 847 MEOGTSDVSGAMEEBAKLEGISPMALTOEGEGTSTPLVEELNLPAMKKDPRESSGRK 906

DB 265 RENGSVESDGLTNDSSSLMGDOEYQSRSPDILETTTSFOALSPANQAESTIKSPDAGSK 324

XX 907 A 907

DB 325 A 325

Query Match	Best Local Similarity	Matches 128	Conservative	42	Mismatches 128	Indels 123	Gaps 13
510	LMKAVEPKSKADETNP-----	PGSEGS-ALAGVADSGSATRMQLSKLVTSLPSWALL	560				
5	LSSTGNRKDLGSLPDLGRCPRSPESGCTPLRGVGNPNNSPR-----		48				
561	TNHLKSTGSPFPFVYVLEBGLASPSSETSKLQDLVEKIDROGAVAVASTASGAPTTAPAFS	620					
49	-----AGGFGSGSTGPRP-----GSETTKLQDLVENDIK-----		76				
621	SSASGRPOCVYLCILVLSGPRALRLHYOGHGERPRPKCVCGRASTGRNLAHVGHKTS	680					
77	ATTPRECELCILCHVLSGQSSLSKMHVYRTHTGERPRQCKICGAFSTKGNLKTHTLVHRTN	135					
681	PAARAQNSCPTCKKFTFNNAVYTLQOHVNRHLGGQILRPNNGSALSEGGAQENSQSTASG	740					
136	TSATKTONSCPTCKKFTFNNAVYTLQOHVNRHLGGQILRPNNGSALSEGGAQENSQSTASG	171					
741	PGSPPOPOSOOPSPFEEMSEEBEEDDEEDVDVDDSLAGSGSEGGGKAI-----SV	793					

Db	172	-----TLPENPC-----DFTGSEPT--VGENGSTGACHDVIEST	207	
Oy	794	RGDSEVSGAEVEEVAITSVAAPTLYKENDSNK----	APQHTLPP--PPPDNLHPOP	846
Db	208	-----DVEEVS--SQEAPSSSSKVPPTPLPSHSASPTLGFAMASIDAGKVPAPFNLORGS	264	
Oy	847	MEQGTEDVSGAMEEAKLEGISPMALTOEGEGSTPLVEELNPEAKRKDGESSGRK	906	
Db	265	RENGSVESDGLTNDSSSLMDQEQVQNSPLLETTSPGALSPANQAEISIKNSPDAGSK	324	
Oy	907	A	907	
Db	325	A	325	

```

RESULT 23
AAM70641
ID AAM70641 standard; Protein; 336 AA.
XX
XX AC
XX AAM70641;
XX
XX DT 06-NOV-2001 (first entry)
XX
XX DE Human bone marrow expressed probe encoded protein SEQ ID NO: 30947.
XX
XX KW Human; bone marrow expressed exon; gene expression analysis; probe;
XX microarray; cancer; leukaemia; lymphoma; myeloma.
XX
XX OS Homo sapiens.
XX
XX PN WO200157276-A2.
XX
XX PD 09-AUG-2001.
XX
XX PF 30-JAN-2001; 2001WO-US000668.
XX
XX PR 04-FEB-2000; 2000US-0180312.
XX PR 26-MAY-2000; 2000US-0207456.
XX PR 30-JUN-2000; 2000US-0608408.
XX PR 03-AUG-2000; 2000US-0632366.
XX PR 21-SEP-2000; 2000US-0234687.
XX PR 27-SEP-2000; 2000US-0236359.
XX PR 04-OCT-2000; 2000GB-0024263.
XX
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX DR WPI; 2001-488900/53.
XX
XX PT Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human bone marrow -
XX
XX PS Example 4; SEQ ID NO: 30947; 658pp + Sequence Listing; English.
XX
XX XX The present invention provides a number of single exon nucleic acid
XX probes which are derived from genomic sequences expressed in the human
XX bone marrow. They can be used to measure gene expression in bone marrow
XX samples, which may enable the improved diagnosis and treatment of cancers
XX such as lymphoma, leukaemia and myeloma. The present sequence is a
XX protein encoded by one of the probes of the invention.
XX
XX SO Sequence 336 AA;
XX
XX Query Match 8.3%; Score 438.5; DB 22; Length 336;
XX Best Local Similarity 30.4%; Pred. No. 1.5e-19;
XX Matches 128; Conservative 42; Mismatches 126; Indels 123; Gaps 13;
OY 510 LMKAYEPKSKADENTP-----PGSRGS-AIAGVADSGSARQMQLSKLYTSLPSMALL 560
Db 5 LSSGNGPKDLTGSLPDDLPQGPSPESGSGPTLPVGPVNTSPR----- 48

```


CC The present invention relates to human single exon nucleic acid probes
CC (SENAP: see AA110068-A128459). The present sequence is a peptide encoded
CC by one such probe. The SENAPs are derived from human HeLa cells. The SENAPs
CC can be used to produce a single exon microarray, which can be used for

Query Match	8.3%	Score 438.5	DB 22	Length 336
Best Local Similarity	30.4%	Pred. No. 1.5e-19		
Matches 128	Conservative 42	Mismatches 128	Indels 123	Gaps 13
QY	510	LKKAVPEKSKADENTP-----PGSEGS-AIAGVADSGSATRMQLSLVTSIPSWALL	560	
Db	5	LSSGNTPKDLTGSGSLPEDDLOGPSPESBGPTLPVGGPNVNSPR-----	48	
QY	561	TNHLKSTGSEFPFVYLEPLGASPSSETSKLOOLVEKIDROGAANAFASTASCAPTTSAPAPS	620	
Db	49	-----AGFGSGSCTPEP---GSETTLKLOOLVENIDK-----	76	
QY	621	SSASGNGOCYICLRVLSCPRALRLHYHGGRPFCKYCGRAFSTGNLRAPVGHKTS	680	
Db	77	-ATTDPRECIICHRYLSCQSSLKMHYRTHGTGPPOCKICGRAFSIKGNLKTHLGVHRTN	135	
QY	681	PARAARNSCFICQKKTFTNAVTLQOHVRMHLGQIPNGSALSEGGAOENSSBOSTASG	740	
Db	136	TSIKTHQHSCHRICQKFTNAVMLQOHIRMHMGQIPN-----	171	
QY	741	PGSFPPOSOQPSPEEEMSEEEDEEEDYVTDSESLAGRSGESGGEKAT-----SV	793	
Db	172	-----TLPENPC-----DETGSEPMT--VGENGSTGAICHDDVIESI	207	
QY	794	RGDSEEVSGAEEEVATSVAAPTTVKEMDSNEK---APQHTLPP---PPPDNLDPQP	846	
Db	208	--DVEEVS-SQEARSSSKVPTPLPSIHASPTLGFAMMASLDAFGKVPAPFNLORGS	264	
QY	847	MEDGTSVDSGAMEEBAKLEGJSSPMALTOEGEGTSTPLVEELNPPAMKKDPGESSGRK	906	
Db	265	RENGSVESDGLTINDSSSLMGDEYOSRSPDILETTTSFQALSPANSSQAESIKSPDAGSK	324	
QY	907	A 907		
Db	325	A 325		
RESULT 25				
AAM30956				
ID	AAM30956	standard; Protein; 336 AA.		
XX				
AC	AAM30956;			
XX				
DT	17-OCT-2001	(first entry)		
DE	Peptide #4993 encoded by probe for measuring placental gene expression.			
XX				
KW	Probe: microarray; human; placenta; antenatal diagnosis;			
XX				
OS	genetic disorder.			
XX				
PN	Homo sapiens.			
XX				
XX	WO200157272-A2.			
PD				
XX	09-AUG-2001.			
XX				
XX	30-JAN-2001; 2001WO-US00663.			
XX				
PR	04-FEB-2000; 2000US-0180312.			
XX	26-MAY-2000; 2000US-0207456.			
PR	30-JUN-2000; 2000US-0608408.			
XX	03-AUG-2000; 2000US-0632366.			
PR	21-SEP-2000; 2000US-0234687.			

Db 265 RENGSVESDGLTNDSSILMGDOEYQSRPDILETTTSQLSPANSQAESIKSPDAGSK 324
QY 907 A 907
325 A 325
RESULT 27
ABG40328 standard: Peptide: 336 AA.
ID ABG40328
AC ABG40328;
DT 19-AUG-2002 (first entry)
DE Human peptide encoded by genome-derived single exon probe SEQ ID 29993.
XX Human: single exon probe; asthma; lung cancer; COPD; ILD;
KM chronic obstructive pulmonary disease; interstitial lung disease;
KW familial idiopathic pulmonary fibrosis; neurofibromatosis;
KW tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;
KW Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemosiderosis;
KW pulmonary histiocytosis; lymphangioleiomyomatosis; Karagenen syndrome;
KW pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;
KW primary ciliary dyskinesia; pulmonary hypertension;
KW hyaline membrane disease.
XX Homo sapiens.
OS WO200186003-A2.
PN 15-NOV-2001.
PD 30-JAN-2001; 2001WO-US00665.
PE 04-FEB-2000; 2000US-180312P.
PR 26-MAY-2000; 2000US-207456P.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632386.
PR 21-SEP-2000; 2000US-234687P.
PR 27-SEP-2000; 2000US-236359P.
PR 04-OCT-2000; 2000GB-0024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
PA Penn SG, Hanzel DK, Chen W, Rank DR;
PI WPI: 2002-114183/15.
PT Spatially-addressable set of single exon nucleic acid probes, used to
PT measure gene expression in human lung samples -
PS Claim 27; SEQ ID NO 29993; 634pp; English.
XX The invention relates to a spatially-addressable set of single exon
CC nucleic acid probes for measuring gene expression in a sample derived
CC from human lung comprising single exon nucleic acid probes having one of
CC 12614 nucleic acid sequences mentioned in the specification, or their
CC complements for the 12387 open reading frames derived from the 12614
CC probes. Also included are a microarray comprising the novel set of
CC probes; the novel set of probes which hybridise at high stringency to a
CC nucleic acid expressed in the human lung; measuring gene expression in a
CC sample derived from human lung, comprising (a) contacting the array with
CC a collection of detectably labeled nucleic acids derived from human lung
CC mRNA, and (b) measuring the label detectably bound to each probe of
CC the array; identifying exons in a eukaryotic genome, comprising
CC (a) algorithmically predicting at least one exon from genomic sequences
CC of the eukaryote; and (b) detecting specific hybridisation of detectably
CC labeled nucleic acids from eukaryote lung mRNA, to a single exon probe,
CC having a fragment identical to the predicted exon, the probe is included
CC in the above mentioned microarray, assigning exons to a single gene,
CC comprising (a) identifying exons from genomic sequence by the method
CC above and (b) measuring the expression of each of the exons in several

CC tissues and/or cell types using hybridisation to a single exon
CC microarrays having a probe with the exon, where a common pattern of
CC expression of the exons in the tissues and/or cell types indicates that
CC the exons should be assigned to a single gene; a peptide comprising one
CC of 12011 sequences, mentioned in the specification, or encoded by the
CC probes/open reading frames (ORF). The probes are used for gene
CC expression analysis, and for identifying exons in a gene, particularly
CC using human lung derived mRNA and for the study of lung diseases
CC such as asthma, lung cancer, chronic obstructive pulmonary disease
CC (COPD), interstitial lung disease (ILD), familial idiopathic pulmonary
CC fibrosis, neurofibromatosis, tuberous sclerosis, Gaucher's disease,
CC Niemann-Pick disease, Hermansky-Pudlak syndrome, sarcoidosis, pulmonary
CC haemosiderosis, pulmonary histiocytosis, lymphangioleiomyomatosis,
CC pulmonary alveolar proteinosis, Karagenen syndrome, fibrocystic
CC pulmonary dysplasia, primary ciliary dyskinesia, pulmonary hypertension
CC and hyaline membrane disease. The present sequence is a peptide/protein
CC encoded by a single exon probe of the invention.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX Sequence 336 AA:
SQ
QY 510 LMKAVEPKSKADENTP-----PGSEGS-AIACVADSGSATRMQLSKVTSLPSMALL 560
Db 5 LSSGTNPKDLTGGSLPGDLOGPSPSESGPPLPGVGPYNSPR----- 48
QY 561 TNLKSTGSPFPYVLEPLGASPSETSKLOQLVEKIDGAGAVAVASTAGAPPTSABAPS 620
Db 49 -----AGFGGSGTPEP-----GSETIKLOLVENIDK----- 76
QY 621 SSASGPNQCVICLRVLSCPRALRLHYOGHGERPKCKYCGRAFSTRGNLRAHFVGKTS 680
Db 77 -ATTDPNKECLICHRVLSGSSLSKMHYRTHTGERPFCKICGRAFSFGNLTGLGVRHTN 135
QY 681 PAARAQNSCPTCKOKKNTAAVTLQOHVBMHGGIIPNGSALISGGGAQENSBSQSTASG 740
Db 136 TSITQHSCTPCCKKFTNAVMVLMQHIRMAMGGIIPN----- 171
QY 741 PGSPFPQSQOPSPSEEMSEEEDEEDVDYDSDSLAGSGESGGEKAI-----SV 793
Db 172 -----TPLEPNPC-----DFTGSEPMT--VGENGSTGAICHDDVIESI 207
QY 794 RGDSEEVSGAEDEVATVAAPTTVKEMDSNEK-----APQHTLPPP-----PPPDNDHQP 846
Db 208 --DVEEVS--SSEAPSSSSKVPPTPLPSIHSAPTLGFAMMASLDAIPKVGPAFPFNLRQGS 264
QY 847 MEGCTSDVSGAMEEAKLEGISSPMNALNQEGECTSPVLEELNLPEAMKKDPGESSGRK 906
Db 265 RENGSVESDGLTNDSSILMGDOEYQSRPDILETTTSQLSPANSQAESIKSPDAGSK 324
QY 907 A 907
Db 325 A 325
RESULT 28
AAM80283
ID AAM80283 standard: Protein: 927 AA.
XX
AC AAM80283;
DT 06-NOV-2001 (first entry)
DE Human protein SEQ ID NO 3929.
XX Human: cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;

PF 28-JUL-2000; 2000EP-0116126.
XX 29-JUL-1999; 99JP-0248036.
PR 27-AUG-1999; 99JP-0300253.
PI 11-JAN-2000; 2000JP-011876.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX (HELI-) HELIX RES INST.
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
DR WPI: 2001-318749/34.
XX Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
PS Claim 8: SEQ ID 17076; 2537Pp + CD ROM; English.
XX The present invention describes primer sets for synthesizing 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-qt primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to 'e' and
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to
CC AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX
XX Sequence 619 AA:
SQ
Query Match 6.2%; Score 330.5; DB 22; Length 619;
Best Local Similarity 21.4%; Pred. No. 1.8e-12;
Matches 173; Conservative 78; Mismatches 276; Indels 281; Gaps 34;
OY 218 PASPSLPGTGAASSTKPLPLSPFIKAQOTGKTATSSSSSSSSGAEP-----PKQ 269
DB 46 PADCSSE-----EVAEVRK-----KPETEAKAEASGCEKAVGSAKPPRYACPLCEKA 92
OY 270 AFHLYLPLGSHQHFVSGVGSRH---KTPAPSPALPGSTDLLIAP-----HLAEPGT 321
DB 93 ---YKTAPELRSH-----GRSHGKEKPPCP-----ECGRKFMQVCLRHVLA----- 132
OY 322 TGLLAOACIGAARGLEAASAPGLRKPKNSGELGYEVISSLEKPGGRHKRCFCAKVGES 381
DB 133 -----SHAGELRP-----RAHCCKAYGA 151
OY 382 DSALQIHLKRSHGTGERPYCANNCGNRTTGNLKVHRRHREKPYHVOMNPHVPEHLLYV 441
DB 152 LSKLKIHQRGHTGERPYACADCGKSPADS---VFRKHRR--THAGLRP-----YS 197
OY 442 ITSSGLPGMSVPPPKAEBEACTPGGCVERRKPLVAST--TALSTATSTLLSTGTSTAVA 499
DB 198 CERGGKAYA-ELKDLRNHRSHIT-----GERPFLSCGKGSFSRSSSLTCHQRTIAAQKP 251
OY 500 PGLPTFNK-FVLMKAVEPKSADENTPPGSEGSAINAGVADSGSATRMQLSKLVYSLPSWA 558

DB 252 YNCPACGKGFTQLSSVQSHERTHSGEKPELCPRCGRMFSDPSSFRHQRAR----- 302
OY 559 LITNLKSTGSPFPFVLEPLG---ASPSETSKLOQLEYEKIDRGAVNAVASTSGAPTTS 615
DB 303 ---EGVK-----PYHCEKCGKDFROPADLA-MHRVHTGDR----- 334
OY 616 AAPASSASGPNOCVICRLVYLCPRALRLHYGOHGERPFKCKVCGRAVSTRGNLRAHFV 675
DB 335 -----PFKLCQCDKFTFVASMCLKRHIALVHSGQRFRCRECGRAFAERASLTKRHSR 384
OY 676 GHKTSAPAARAONSCTPQCKFTNAVTLQOHWRLGQ---IPNG-----GSLSBEGGA 727
DB 385 VH-----SGERPFHCNACGKSPVSVSSILRKHERTHRSSEAGVPPAOLVYGLALPV--GV 438
OY 728 AOPNSSEQSTASGPGSFPPQSQSPPEEKSEEEDEDEEDVDYDDESLAGRSSEGG 787
DB 439 AGSSSAAPAGAGIGDPPA-----GLLGLPPSSGG 468
OY 788 EKALSVRGDSEEVSGAEAEVATSVNAPTVKEMDSNEKAPQHTLPPPPPPDNLHPQM 847
DB 469 -----VMATQMQVYGM-----TVEHVECO 487
OY 848 EQGTSVSGAMEEERAKLEGISSPMALTOEGEGSTPLVYELNLPAMKADPGESSGRKA 907
DB 488 DAGVREAPGPV-----EGAG-----EAGGEADEKPPQFV 517
OY 908 CEVCGGSPFPQTALENHOKTKPKDGLPTCVFCRQGLDRLATLKKHMLNHOVPPF-AP 966
DB 518 CRCKETFFSTTLRLRHRSHHELRP-PPCTQCGKISIDRAGLRKHS-RTHSSVPYTCP 575
OY 967 HGPONITLSLVPCSSSIPSGLSPPF 994
DB 576 HCKPAFLSADLRKHERRHPVMTPTP 603
XX
XX RESULT 30
XX AAM79299
XX ID AAM79299 standard; Protein; 869 AA.
XX AC AAM79299;
XX DT 06-NOV-2001 (first entry)
XX DE Human protein seq ID NO 1961.
XX DE Human protein seq ID NO 1961.
XX KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorder; arthritis; inflammation.
XX Homo sapiens.
XX WO200157190-A2.
XX 09-AUG-2001.
XX 05-FEB-2001; 2001WO-US04098.
XX 03-FEB-2000; 2000US-0496914.
XX 27-APR-2000; 2000US-0560875.
XX 20-JUN-2000; 2000US-0598075.
XX 19-JUL-2000; 2000US-0620325.
XX 01-SEP-2000; 2000US-0654936.
XX 15-SEP-2000; 2000US-0693561.
XX 20-OCT-2000; 2000US-0693325.
XX 30-NOV-2000; 2000US-0728422.
XX
XX (HYSE-) HYSEQ INC.
XX Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;
PI Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;

SO	Sequence	702 AA;			
Query Match	6.1%;	Score 322.5;	DB 23;	Length 702;	
Best Local Similarity	20.3%;	Pred. No. 6.7e-12;			
Matches 156;	Conservative 71;	Mismatches 287;	Indels 263;	Gaps 23;	
OY	231	SSTKPLPLFSPRIKPAOTGKTWASSSSSSSSSSGAEPPKQAFHLIYHPLGSHPFESVGWG	290		
DB	122	TDTLPR-----KPKCKNEKGKIIISDGSNOKRLPEKRP-----HPCGECGGG	161		
OY	291	RSKRPTRPASPALPESTDOLIASPHILAFEGTTGLLAQCSGANGLEAASBGLIKPK--	348		
DB	162	FSY-----SRLLPLRPN--VHTGKCFSSSHLRTHQRIHPGKELNCHESGDCFNKSS	213		
OY	349	-----NGSGELGYGEVVISLSLEKPGGRHKRCFCAKVFGSDSALDIHLRSHNGERYKCV	402		
DB	214	PHYSQSNHTGKESY-----RCHSGCKGFSSSTGLIITHYTHGTEKRYKCEE	259		
OY	403	CGNBFTRGNLKVHFHNRKERYHYQMNPRVPEHLIDYVITSSGLPYGMSV-----	453		
DB	260	CGKCFSSSNFCQCHRVNTEKRP-----YKCECGCKGFGMSVYLRYHQRVN	305		
OY	454	--DPEKAEEGAAGTGGGVERKRPVLAATALESLLTSTGTAVALRPTENKFEVL	510		
DB	306	RGEKRYKCEE-----CGKGFYQAANFHI	328		
OY	511	MKAV--EPKSKADENTPPGSEGSALAGVADSGSATRMOLSKLVTSLPSMALITNHLKST	567		
DB	329	HQRNHTGKRYKCD-----VCGKGFSSNS-----PLICHRRVHT	362		
OY	568	GSFPF-----PYLEPLGASPSSETSKLQOLEKIDROGA	601		
DB	363	GEKRYKCEACGCKGFRNTDLNHFVRNHTGKRYKCEGCKGFSQASNLQ--VHQNVHTE	420		
OY	602	VAVASTAGARTTAPAPASS-----SASGPNQCVTLKRVLSCPRALRLHYGONGERPRK	656		
DB	421	KRFCECTGCKGFSQSSKLTQTHQVHNTGKERYRCDYVCGKGFSSYMLKLHQVYHNTGKERYK	480		
OY	657	CKVGRASTBRLNLRHNFHGHTKSPARAQNSCRICQKFFNAVTLQOQHVHNLGQIQIN	716		
DB	481	CEEGCKGFSWNRNLNHNQVH-----GSEKRYKCEQCDKFSQALIDFRVHQVHNTGKERYK	536		
OY	717	GGSALSEGGAQENSSSEOSTASGSPFPOQSPSEEMSEEEDEEDEDVTDDED	776		
DB	537	CGVC--GKGFSSSGLQSHQVHNTGKRY-----KC	565		
OY	777	SLAGRSESGEKAISVRQDSEEVSGAEVEAVTSVAAPTVMKENDSNEKARQHTLPPRRP	836		
DB	566	DVSCGKGFYSSQFIYHQRGHTGKERYKCEGCKGFGKRSLSL-----	604		
OY	837	PRDNLDPORPMOGTSDVSGAMEEELKLEGISPMALITGSEGTSTPLVEELNPREAMK	896		
DB	605	---NLRNHQRVH--TGEKPHICEGCG-----APSLPSNLR	635		
OY	897	KDPGSESGRR--ACEVCGSFPQTQALAEHQKTHPKDPLFTVCVFCROGFLDRATLKKH	953		
DB	636	VHLGVHTREKRLKCEGCKGFGSASARLEAHQVHNTGKRP-YKCDICDADFRRRSRLTYH	693		
RESULT 32					
AAB94669					
ID	AAB94669	standard; Protein; 498 AA.			
AC	AAB94669;				
DT	26-JUN-2001	(first entry)			
XX					
DE		Human protein sequence SPO ID NO:15609.			
XX					
KW		Human; primer; detection; diagnosis; antisense therapy; gene therapy.			
OS	Homo sapiens.				
XX					

PN	EPI074617-A2.				
XX					
PD	07-FEB-2001.				
XX					
PE	28-JUL-2000; 2000EP-0116126.				
XX					
PR	29-JUL-1999; 99JP-0248036.				
PR	27-AUG-1999; 99JP-0300253.				
PR	11-JAN-2000; 2000JP-0118776.				
PR	02-MAY-2000; 2000JP-0183767.				
PR	09-JUN-2000; 2000JP-0241899.				
XX					
PA	(HELI-) HELIX RES INST.				
XX					
PI	Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;				
PI	Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;				
XX					
DR	WPI; 2001-318749/34.				
XX					
PT	Primer sets for synthesizing polynucleotides, particularly the 5602				
PT	full-length cDNAs defined in the specification, and for the detection				
PT	and/or diagnosis of the abnormality of the proteins encoded by the				
PT	full-length cDNAs -				
XX					
PS	Claim 8; SEQ ID 15609; 2537pp + CD ROM; English.				
XX					
CC	The present invention describes primer sets for synthesizing 5602				
CC	full-length cDNAs defined in the specification. Where a primer set				
CC	comprises: (a) an oligo-dT primer and an oligonucleotide complementary				
CC	to the complementary strand of a polynucleotide which comprises one of				
CC	the 5602 nucleotide sequences defined in the specification, where the				
CC	oligonucleotide comprises at least 15 nucleotides; or (b) a combination				
CC	of an oligonucleotide comprising a sequence complementary to the				
CC	complementary strand of a polynucleotide which comprises a 5'-end				
CC	sequence and an oligonucleotide comprising a sequence complementary to a				
CC	polynucleotide which comprises a 3'-end sequence, where the				
CC	oligonucleotide comprises at least 15 nucleotides and the combination of				
CC	the 5'-end sequence/3'-end sequence is selected from those defined in				
CC	the specification. The primer sets can be used in antisense therapy and				
CC	in gene therapy. The primers are useful for synthesizing polynucleotides,				
CC	particularly full-length cDNAs. The primers are also useful for the				
CC	detection and/or diagnosis of the abnormality of the proteins encoded by				
CC	the full-length cDNAs. The primers allow obtaining of the full-length				
CC	cDNAs easily without any specialised methods. AAH03166 to AAH13628 and				
CC	AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to				
CC	AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632				
CC	represent oligonucleotides, all of which are used in the exemplification				
CC	of the present invention.				
XX					
SO	Sequence 498 AA;				
Query Match	6.1%;	Score 321.5;	DB 22;	Length 498;	
Best Local Similarity	22.1%;	Pred. No. 5e-12;			
Matches 141;	Conservative 70;	Mismatches 237;	Indels 191;	Gaps 24;	
OY	371	KCRFCAKVFGSDSALQIHLRSHNGERYKCVNCGRRFTTRGMLKVFNHNRKXYHYQMN	430		
DB	20	KCAHCKRAIGALSRLKLNQHGHTGERPRACADGCSFADPS-----VFRKHRR--TAAGLR	73		
OY	431	PRVPEHLIDYVITSSGLPYGMSVPEKAEEDAGTGGVERKPLVAST--TALSTATESLT	488		
DB	74	P-----YSCERCGKAYVA-ELKDLRNHERSHT-----GERPFLCSECGKFSRSSSLT	119		
OY	489	LSTGTSTAVAPGLPTTNK-FYLMKAVPRKSKADENTPPGSGSALAGVADSSGATPMQL	547		
DB	120	CHQRIHAQKPYRCBPACGKGFQTLSSYQSHERTHSGEKPFILCPRCGRMFSDPSFRRHQR	179		
OY	548	SKLVTSLSWALLTNHLKSTGSPFPYVLEPLG--ASPSETSKLQOLEKIDROGAVAL	604		
DB	180	AH-----EGVK-----PYHCEKCGKPRORPADLA-MHRRVNTGGR-----	213		
OY	605	ASTASGAPTTAPAPSSSASGPNQCVICLRYLSCPRALRLHYGONGERPFKCYGCGRAF	664		

Query Match	5.98;	Score 311.5;	DB 22;	Length 744;
Best Local Similarity	18.18;	Pred. NO. 3.5e-11;		
Matches 193;	Conservative 90;	Mismatches 281;	Indels 505;	Gaps 37;
90	MDTEHSNPDDSSSGPPDPTWGPERRGESESGOFLVA-ATGTAAAGGGGGLILASPKLG--	146		
1	MTPESSSSSEISGGGCAITPLRPSR-----MDQFNMSKMAAAAAAGGGGLPGAADRNNGS	55		
147	-----ATPLPPESTR-----APPPEPP-----	164		
56	GGSDGSGONGNGDSRRNSSASRISAYETOLAYQOHLGLGHGPPPPPPSHHREISAFVPVL	115		
165	---PPPPGVGSHLNPILILEELRVLDQORQIHQOMTEQICROVLLIGSGQYGAAPSP	221		
116	PTGKVRPSNSNYETIITAMADKKRELRLR-----AAAAAAMGGRGGGPGGP	163		

[illegible]

QY	205	VLLGSLGQYCAPARSP--SELRPGCAASTRYLPLFSPKRAQOIGKTTAASSSSSSSS	262
Db	569	NALGGLRPLGLASSSPPLHNEERDEEHODEDL-----EOKERLASQSED----	614
QY	263	GAEPKQAFHLYHPLGSONHPFVSGVGRSHKPTP-----APSP	301
Db	615	-----FFHOLYLKHKTANCGALSHPPSPKIKHEPADOKDLADIOSILNMTSSSS	663
QY	302	ALPGSTDOLIASPH--LAFPGTTGLLAQCSLGA--RG-----LEAASPGLK	346
Db	664	SFLRNFEOGSVNPNSQOYSLDRDDEEEOADAFTESEFRMKLGEFPCKLCTAVFPMILRA	723
QY	347	PKNGSGELGYGVISLLEKPRGRHKRCFCAKAVGSDSALOIHLSHSGEPRKYCNWGNR	406
Db	724	LK-----GHNVHLIGA VGRPA GFRKCMCPYAVCSDAALYRHRMTHNGDRPYEACAVCNYA	777
QY	407	FTTRGNLKVHF-HRH--REK-----YPHVOMNP	431
Db	778	FTTKANCSEHILRNHRIGKTSREEVKRAIYVYHRAEDACCEDSKSRLEDLADTSFBSISPTP	837
QY	432	HPVP-----EHLDVYIIS--GLRPGMSVPRPKAE	459
Db	838	PPRPVNESKSQLKHMILGENHLAPVMOQRPPLKIQVANSLOLVYKKBAPARPQOQOQOQOQ	897
QY	460	BEAGT-----PGGVEVERKPIVASTTALASATSELTLLTSGTSTAVAPGLPTF	505
Db	898	EKSSGALDRPSMDVLDSLKKPTGSGASLTIPAVTRP-----TPAAVAP-----	938
QY	506	NKFVLMKAVERKSKADENTPPGSEGSALAGVADSGSATRMO-----LSK	549
Db	939	-----YPRGG-----VTPRIALAAIEBOOILLAAOOOLFAGAGEYMOQ	975
QY	550	LVTSLPSMALLTNHLKSTGSPF-FPVVLEPLGASPEFTSKLOOLVEKIDRQGAVAVASTA	608
Db	976	LFRSL-----MFGSQTSGFRFFPRPMARP--PROANPKRPVWSPPNKRNPMVP--GVG	1024
QY	609	SGARTTSAPARSSASGRNQCYICLVLSCPRALRLHYGONGGERPKVCYCGRAFSTRG	668
Db	1025	VGVVP-----PRGCPYKMYIKNGVL--MPKOKORYRT--ERPRACEHSGARFTLRS	1072
QY	669	NLRAPF-----VGKHTSPAARQON-----	687
Db	1073	NMRHVKQOHPOFYAORORSNGHHVMRGSGASNVAAAAAATAATVMAGGPSSGFGS	1132
QY	688	-----SCPICQ-----KFTNAVYLOQHV-----MHL	710
Db	1133	NHHNGHGSHSGHGNARPISEYQKCALMLOQLKANKTDLLOQALHAGSSSVACNPRLLHF	1192
QY	711	GGQIRP-----GGSALISEGGGA-----QENSSSEDST--ASGPGSPRPOQSO	750
Db	1193	GPRLPNPSRPMHNGS--SQNGOQATTAMDDERPKLIDEDENENHVEVAEDVDF--E	1245
QY	751	OPSRSEEESEEE-----DEEEEDVTDSDS-----LAQRSGSGGE-----	788
Db	1246	EDEDEEMERDEPERLIDDEQRAKKEABEEOELPKLEBQLGTEAAQKMAETLLEQAIK	1305
QY	789	--KAISVRGDSEEVGAEEBVAAT-----SVAAPTT-----VKEMDSN	823
Db	1306	ACKRPLSRPTKEMASPA NPTVAATMOERPAITAASTNSSLKTMIQAQEVYKSLKEVASS	1365
QY	824	E-KARPHTLRPPRRPNLDHPRMEQST--SDVSGAMEEELKLGISIPMALPDQEG	878
Db	1366	PRKDESDQLVPAKLYDNAT--SONMGFNSTFRSDVANHHQOSE--EOLVASGSSESNN	1422
QY	879	EGSTPRVLEELNLRPMKKKDPGESSGKACVGSQSPFTTALAEHOKTHPKDPRLFYCV	938
Db	1424	SGTEDVTSSSSSEPKKSAVSLAPRNVSCPYO RMPFWSSSLRHLITHTGQRP-FKCS	1482
QY	939	PCRQGLDRATLKKHMLLHNHOV	961
Db	1483	HCRLFTTYSNCDRIHLRKHGV	1505

	RESULT	39
XX	ABG28743	
XX	ABG28743 standard; Protein; 1196 AA.	
AC		
XX	ABG28743;	
DT	18-FEB-2002 (first entry)	
DE	Novel human diagnostic protein #28734.	
XX		
KW	Human; chromosome mapping; gene mapping; gene therapy; forensic;	
XX	food supplement; medical imaging; diagnostic; genetic disorder.	
OS	Homo sapiens.	
PM	WO200175067-A2.	
PD	11-OCT-2001.	
XX		
PF	30-MAR-2001; 2001MO-US08631.	
PR	31-MAR-2000; 2000US-0540217.	
PR	23-AUG-2000; 2000US-0649167.	
XX		
PA	(HYSE-) HYSEQ INC.	
PI	Dymanac RT, Liu C, Tang YT;	
DR	WPI; 2001-639362/73.	
DR	N-PsDB; AAS92930.	
XX		
PT	New isolated polynucleotide and encoded polypeptides, useful in	
PT	diagnostics, forensics, gene mapping, identification of mutations	
PT	responsible for genetic disorders or other traits and to assess	
PT	biodiversity -	
PS	Claim 20; SEQ ID NO 59102; 103pp; English.	
XX		
CC	The invention relates to isolated polynucleotide (I) and	
CC	polypeptide (II) sequences. (I) is useful as hybridisation probes,	
CC	polymerase chain reaction (PCR) primers, oligomers, and for chromosome	
CC	and gene mapping, and in recombinant production of (II). The	
CC	polynucleotides are also used in diagnostics as expressed sequence tags	
CC	for identifying expressed genes. (I) is useful in gene therapy techniques	
CC	to restore normal activity of (II) or to treat disease states involving	
CC	(II). (II) is useful for generating antibodies against it, detecting or	
CC	quantitating a polypeptide in tissue, as molecular weight markers and as	
CC	a food supplement. (II) and its binding partners are useful in medical	
CC	imaging of sites expressing (II). (I) and (II) are useful for treating	
CC	disorders involving aberrant protein expression or biological activity.	
CC	The polypeptide and polynucleotide sequences have applications in	
CC	diagnostics, forensics, gene mapping, identification of mutations	
CC	responsible for genetic disorders or other traits to assess biodiversity	
CC	and to produce other types of data and products dependent on DNA and	
CC	amino acid sequences. ABG00010-ABG30377 represent novel human	
CC	diagnostic amino acid sequences of the invention.	
CC	Note: The sequence data for this patent did not appear in the printed	
CC	specification, but was obtained in electronic format directly from WIPO	
CC	at ftp.wipo.int/pub/published_pct_sequences.	
XX		
XX		
SQ	Sequence 1196 AA;	
	Query Match 5.8%; Score 305.5; DB 22; Length 1196;	
	Best Local Similarity 21.3%; Pred. No. 1.5e-10;	
	Matches 143; Conservative 72; Mismatches 238; Indels 219; Gaps 23;	
OY	346 KPKNGSGELGEVYISLEKPGGRKCRFCACVFGSDAQLHLRSHTGEPKYKCNCGN 405	
Db	I I I I I I I I : : : : I I I I : : I I I I I I I I I I	
	358 KATNLSSPFGHEHITHSAKP---YECKCGKAFRASQLRVHGHTGTGKPEYCKRCGK 414	
OY	406 RFTTGNLKVHFHRHREKYP-----HVQMNPHPVEPHLDIVYTSSGLPYGM 451	
	I : : I I I I I : : I : : I I I I I I I I I I	

Db 426 -----CGSFFORRSSLKRLH-RIHARDKDRRSSESGS-----RRDSDRRPFV 468
QY 629 CVICLVLSQFRLRLHLYGOGHGERPFCKXVCGRAFTSTGNLRAHFGHKTSPAARQNS 688
Db 469 CSDCGKAFRSEHLVAHRRVHTGERPFSCQACGRSFTQSSQL---VSHQRVHTGEKPYA 524
QY 689 CPICOKFTNAVTLQOHVRMHLGQIPNGSALSSEGGNAQENSSEOSTASGPGSFPOQ 748
Db 525 CPQCGKRFRVRRASLARHLTH-GGPRPHCTQCGKSFQGTODLARHQRSHTG----- 575
QY 749 SQQSPFEEMSEEEDEEEDVTDSDSLAGRSESGGEKAI SVRGDSEEVSGAEFEVA 808
Db 576 -EKPC-----RCSECG-----EGFSQSAHILARHQ-- 598
QY 809 TSVAAPTVYKEMDSNEKAPQHTLPPRRPDLNLDHPQPMEGTSDVSGAMEEAKLEGIS 868
Db 599 -----RIHTGEK-----PHACDTGCH-----RFRNSS 620
QY 869 SPMAALTOEGEGTSTPLVEELNLPKAMKDPGESSGRKACEVCGSFTOTALLEHOKTH 928
Db 621 -----NLARHRSHTGERP--YSCQTCGRSFRRAHILRHLATH 657
QY 929 PKDG-----PLFTCVFCRQGFLDRA TLKKNMLLAHHQVPPFA 965
Db 658 AEPGOEQABEPPOECVECGKSFSCNLLRH-LVHTGARPY 698

Search completed: January 13, 2003, 15:20:09
Job time : 54.9432 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 13, 2003, 15:17:45 ; Search time 14.9776 Seconds
(without alignments)
1668.393 Million cell updates/sec

Title: US-09-988-117-3

Perfect score: 5301

Sequence: 1 MAQDTGSSRLGPGCEPAE.....SSITPGLSPFRKDDPTMP 1002

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/1/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/1/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/1/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/1/1aa/PCTUS.COMB.pep:*
6: /cgn2_6/ptodata/1/1aa/backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	290	5.5	1185	4	US-09-041-886-23
2	260	4.9	543	4	US-09-362-123A-4
3	255.5	4.8	803	4	US-09-063-035-2
4	249.5	4.7	1706	2	US-08-459-568-2
5	249.5	4.7	1706	2	US-08-399-411-2
6	249.5	4.7	1706	3	US-08-516-859A-2
7	249.5	4.7	1706	4	US-09-586-472-2
8	249.5	4.7	1706	4	US-09-528-706-2
9	236	4.5	1719	2	US-08-459-568-4
10	236	4.5	1719	2	US-08-399-411-4
11	236	4.5	1719	3	US-08-516-859A-4
12	236	4.5	1719	4	US-09-586-472-4
13	236	4.5	1719	4	US-09-528-706-4
14	234.5	4.4	640	4	US-09-262-773-4
15	234.5	4.4	648	4	US-09-262-773-2
16	228.5	4.3	728	5	US-08-475-844-5
17	228.5	4.3	728	5	PCT-US95-08429-5
18	226	4.3	533	1	US-08-040-548-1
19	226	4.3	533	1	US-08-466-344-1
20	226	4.3	533	6	5206152-2
21	226	4.3	3969	4	US-08-061-376-5
22	225.5	4.3	727	2	US-08-475-844-9
23	225.5	4.3	727	5	PCT-US95-08429-9
24	224.5	4.2	543	2	US-08-224-482-4
25	224.5	4.2	2972	4	US-09-579-181-2
26	224.5	4.2	3118	4	US-09-579-181-1
27	222	4.2	449	1	US-08-102-942A-4

28	222	4.2	449	4	US-09-037-179B-4	Sequence 4, App11
29	220.5	4.2	496	2	US-08-224-482-2	Sequence 2, App11
30	217	4.1	429	1	US-08-234-783-4	Sequence 4, App11
31	217	4.1	429	1	US-08-456-907-4	Sequence 4, App11
32	217	4.1	429	5	PCT-US95-05523-4	Sequence 4, App11
33	216.5	4.1	510	4	US-08-246-489-2	Sequence 2, App11
34	215.5	4.1	449	1	US-08-102-942A-6	Sequence 6, App11
35	215.5	4.1	449	4	US-09-037-179B-6	Sequence 6, App11
36	211.5	4.0	1093	5	US-08-545-860D-55	Sequence 55, App1
37	211.5	4.0	1093	5	PCT-US94-04496-55	Sequence 55, App1
38	211.5	4.0	1958	1	US-07-945-283-2	Sequence 2, App11
39	208.5	3.9	671	4	US-09-121-321-16	Sequence 16, App1
40	208.5	3.9	671	4	US-08-933-803A-16	Sequence 16, App1
41	208	3.9	432	4	US-08-711-417C-197	Sequence 197, App
42	204.5	3.9	488	2	US-08-933-750C-17	Sequence 17, App1
43	204.5	3.9	488	4	US-09-224-613-17	Sequence 17, App1
44	204	3.8	288	4	US-09-432-985-2	Sequence 2, App11
45	204	3.8	1135	2	US-08-574-959A-7	Sequence 7, App11
46	204	3.8	1135	4	US-09-357-014-7	Sequence 7, App11
47	203	3.8	461	4	US-08-711-417C-196	Sequence 196, App
48	203	3.8	516	4	US-08-711-417C-202	Sequence 202, App
49	200	3.8	905	2	US-08-574-959A-9	Sequence 9, App11
50	200	3.8	905	4	US-09-357-014-9	Sequence 9, App11
51	200	3.8	1291	4	US-09-150-460B-10	Sequence 10, App1
52	200	3.8	1291	4	US-09-220-641-5	Sequence 5, App11
53	195	3.7	470	2	US-08-946-241B-2	Sequence 2, App11
54	195	3.7	470	2	US-09-309-053-2	Sequence 2, App11
55	195	3.7	479	2	US-08-946-241B-9	Sequence 9, App11
56	195	3.7	479	3	US-09-309-053-9	Sequence 9, App11
57	195	3.7	547	1	US-08-340-203A-3	Sequence 3, App11
58	195	3.7	547	2	US-08-452-567-3	Sequence 3, App11
59	195	3.7	547	2	US-08-452-567-3	Sequence 3, App11
60	195	3.7	547	3	US-09-085-407-3	Sequence 3, App11
61	195	3.7	711	2	US-08-820-170A-10	Sequence 10, App1
62	195	3.7	711	3	US-09-035-699-10	Sequence 10, App1
63	195	3.7	711	4	US-09-273-565-10	Sequence 10, App1
64	195	3.7	711	4	US-09-565-538-10	Sequence 10, App1
65	195	3.7	711	4	US-09-661-468-10	Sequence 10, App1
66	193.5	3.7	1581	4	US-09-110-517-2	Sequence 21, App1
67	193	3.6	361	4	US-08-874-569B-21	Sequence 21, App1
68	193	3.6	368	5	PCT-US93-08743-5	Sequence 5, App11
69	189.5	3.6	944	4	US-09-449-285A-2	Sequence 2, App11
70	189.5	3.6	976	4	US-08-894-997-50	Sequence 50, App1
71	189.5	3.6	1321	2	US-08-317-310A-64	Sequence 64, App1
72	188.5	3.6	706	1	US-08-074-967-2	Sequence 2, App11
73	188.5	3.6	706	4	US-08-553-541B-2	Sequence 2, App11
74	188.5	3.6	706	4	US-09-268-202-2	Sequence 2, App11
75	188.5	3.6	706	5	PCT-US94-06669-2	Sequence 2, App11
76	187	3.5	455	2	US-08-224-482-6	Sequence 6, App11
77	187	3.5	456	1	US-08-040-548-2	Sequence 2, App11
78	187	3.5	456	1	US-08-466-344-2	Sequence 2, App11
79	186	3.5	376	4	US-08-874-569B-20	Sequence 20, App1
80	184	3.5	2441	3	US-08-961-739-2	Sequence 2, App11
81	183	3.5	518	4	US-08-711-417C-198	Sequence 198, App
82	183	3.5	2441	1	US-08-194-468-2	Sequence 2, App11
83	183	3.5	2441	4	US-09-514-247A-8	Sequence 8, App11
84	182.5	3.4	1312	4	US-09-041-886-19	Sequence 19, App1
85	182.5	3.4	1290	4	US-09-150-460B-6	Sequence 6, App11
86	180.5	3.4	1315	3	US-08-899-595-3	Sequence 3, App11
87	180	3.4	2035	1	US-08-046-585-5	Sequence 5, App11
88	180	3.4	2035	1	US-08-393-703-5	Sequence 5, App11
89	180	3.4	2035	5	PCT-US93-11721-5	Sequence 5, App11
90	179.5	3.4	903	2	US-08-853-310-2	Sequence 2, App11
91	176.5	3.3	387	2	US-08-224-482-8	Sequence 8, App11
92	176.5	3.3	2414	1	US-08-227-536-2	Sequence 2, App11
93	176.5	3.3	2414	5	PCT-US95-04682-2	Sequence 2, App11
94	175.5	3.3	453	6	5206152-7	Sequence 2, App11
95	175.5	3.3	470	2	US-08-465-590-153	Sequence 153, App
96	175.5	3.3	470	4	US-08-283-300A-13	Sequence 13, App1
97	175.5	3.3	470	4	US-08-711-417C-153	Sequence 153, App
98	175.5	3.3	470	5	PCT-US95-09345-13	Sequence 13, App1
99	175	3.3	1234	5	US-08-317-310A-15	Sequence 15, App1
100	175	3.3	1234	2	PCT-US95-13041-15	Sequence 15, App1

686 -----ATADETEVAKAE-----ISKAVQVQEDPMTHT----- 715
 QY 890 NUPKAMKDDGESSGRKACEVCGSGSPPTOTALSEHOKTH 928
 Db 716 -----LYACDSCGDKFLDANSIAOHVRIH 739

RESULT 4

US-08-459-568-2
 Sequence 2, Application US/08459568

Patent No. 5811304

GENERAL INFORMATION:

APPLICANT: Huang, Shi

TITLE OF INVENTION: Retinoblastoma Protein - Interacting

TITLE OF INVENTION: Zinc Finger Proteins

NUMBER OF SEQUENCES: 93

CORRESPONDENCE ADDRESSES:

ADDRESSEE: Campbell and Flores

STREET: 4370 La Jolla Village Drive, Suite 700

CITY: San Diego

STATE: California

COUNTRY: USA

ZIP: 92122

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/459,568

FILING DATE: 02-JUN-1995

CLASSIFICATION: 514

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/399,411

FILING DATE: 06-MAR-1995

ATTORNEY/AGENT INFORMATION:

NAME: Campbell, Cathryn A.

REGISTRATION NUMBER: 31,815

REFERENCE/DOCKET NUMBER: P-LJ 1264

TELECOMMUNICATION INFORMATION:

TELEPHONE: (619) 535-9001

TELEFAX: (619) 535-8949

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 1706 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-459-568-2

Query Match 4.7%; Score 249.5; DB 2; Length 1706;

Best Local Similarity 22.1%; Pred. No. 2e-08;

Matches 186; Conservative 69; Mismatches 281; Indels 307; Gaps 38;

14 PCGEPAERGGDASEEHPOVAK-----CCAQ-----FSDPTEFLAHONSCCTDP 58

851 PCD---SEGKEFKENHLAQPAPAKKKPTCKMLQKVLNEYNGVSLPTE-----TTP 898

59 PWNVIIGQGENPSN-SSASSAPRPE-GHSRSQVMDTEHSNPPDSGSGPPPTWGPERRG 116

899 EV-----TRSPSPCKSPDTPDPDELGPDSGCSVPFAESPPEVVGSPSPPLQT-----A 946

117 EESSGQFLVATGTAAGGGGCLILASPKIGATPLPESTPAPPPPPP-----PPP 167

947 SLSSGQ-----LPLPLPTEPSSPPCPPLYLTVATPPP 979

168 PGVSGSHLNTPLILEELRVLQROQHOMQTEQICRQVLLIGSLGQTVGADAPSGELPQT 227

980 PLLPYVPLSHP-----SSDASPOQCEPSP 1002

228 GAASSTKPLPLPLFS-----PIKP---AQTGKTTASSSSSSSSSSGAEPKQAF 272

Db 1003 FSNTTAQSPPLILSPVSPSPPIPEVEPLMSAASBPPTLSSSSSSSS----- 1052
 QY 273 HLHYPLGSHHPFSVGVGGRSIRKPTPAPSPALP---GSTQQLIAS-PLHAPPGTGLAAQ 328
 Db 1053 ---FPSSS-----CSTSPSPPLSAVSSVSSGNNLEASLAVTK----- 1091
 QY 329 CLGAAGLEAASFGLLPKPKNGSGELGYGEVTSLEKPGGRHRCRCACAFYGSALQIH 388
 Db 1092 -----QESSESEG-LKPKREAPRAGOSVYQ---ETFSKNFLCNVNCESPFSLTKDLTKH 1141
 QY 389 LRSHTGERPYKCNVCGNRFTTRGNLKVH-FHRH-----REKYPHV-QMNPHPVP 435
 Db 1142 LSVHADEWEPKCECVOLFKVKTDLSEHRFLHGVGNIFVCSYCKKREFAFLCNLQOHORD 1201
 QY 436 EHLDYVITSGLGYGMVSP-----PEKAEFEAGTCGGYERKRLVASTFALSATESLTL 490
 Db 1202 LHPDEVCTHHEFSGLRPNQFTDPSKANVEH----- 1233
 QY 491 STGTSTAVALPPLTFNKFVLMKAVEP-KSKADENTPPGSEGSALAGVADSGSATRMOLSK 549
 Db 1234 -----MPSLPE-----EPLTSREELINDSE-----ELYTTIKIMASG 1267
 QY 550 LVTSLPSMAL-LTNHLKSTGSEPPPY-VLEPLGASDETS-----KLOOLV 593
 Db 1268 IRTKDDVRLGLNQHYPSFKPPQYHHRNPMGIGVATNTFTTHNIPQTEFTTARCTCG 1327
 QY 594 EKIDRQG-----AVAVASVAGAPITSAPAPSSASASPNQVICL-----RYLSC 638
 Db 1328 KGVDNMPELKHILACASASAKKRTYPRKNPVLQOTYOPKNGVAVLDNSGKNAFRMRGQ 1387
 QY 639 PRALRLHYGQHGGERPPKCYC-----GRAFSTRGNLRHFVGHKTSPPA 682
 Db 1388 PKRLSFNV-ELGKKMSPKRLKLSALKKKNOVLQKALIQKRAAOKMDLR-----DISEA 1440
 QY 683 ARAQNSCPTQCKKFTNAVTLQOHVBMHLGGQIPNGS-----SALSEGGAQOENSESDSTA 738
 Db 1441 --SSHICPYCDREFTYIGSLNKHAAFAFSCPKPLSPSKRRVSHSKKGHASSSSSDRNS 1498
 QY 739 SGP 741
 Db 1499 CHF 1501

RESULT 5

US-08-399-411-2

Sequence 2, Application US/08399411

Patent No. 5831008

GENERAL INFORMATION:

APPLICANT: Huang, Shi

TITLE OF INVENTION: Retinoblastoma Protein - Interacting

TITLE OF INVENTION: Zinc Finger Proteins

NUMBER OF SEQUENCES: 93

CORRESPONDENCE ADDRESSES:

ADDRESSEE: Campbell and Flores

STREET: 4370 La Jolla Village Drive, Suite 700

CITY: San Diego

STATE: California

COUNTRY: USA

ZIP: 92122

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/399,411

FILING DATE: 06-MAR-1995

CLASSIFICATION: 530

ATTORNEY/AGENT INFORMATION:

NAME: Campbell, Cathryn A.

REGISTRATION NUMBER: 31,815

REFERENCE/DOCKET NUMBER: P-LJ 1264

TELECOMMUNICATION INFORMATION:

1053 ---FPSSS-----CSTSPSPPLSAVSSVSSGDNLEASLPVTFK-----1091
329 CLGAARGLEAAAPGLIKPKNGSGELGYEVISLSLEKPGGRHKCRFCAYKGSALQIH 388
1092 -----QEBSESEG-LKPKKEAPRAGGOSVQ--ETFSKNFICNVCESPLSIDLTKH 1141
389 LRSHTGERPYKCNVCGNRFTRGNLKVH-FHRH-----REKYPHV-QMNPHPVP 435
1142 LSVHAEMWPKCECVQLFKVKTDLSEHRFLHGVGNIFVCYCKKEFAFLCNLQOHORD 1201
436 EHLIDYVITSSGLPYGMSP-----PEKAEAEAGTPGGVVERKPLVASTTALATSESLLL 490
1202 LHPDEVCTHHEFESGTLRPONTDPSKANVEH-----1233
491 STGTSTAVAPGLPTFNKFLVLMKAVEP-KSKADENTPPGSEGSALAGVADSGSATRMOLSK 549
1234 -----MSLPE-----EPLETSREELNDSS-----ELYTTIKIMASG 1267
550 LVTSLSPLSMAL-LTNHLKSTGSPFPFY-VLEPLGASPSSETS-----KLQOLV 593
1268 IKTKDDPVRILGNQHYPSFKPPFQYHHRNPMGIGVATNFTTHNIDQFTTALICTKCG 1327
594 EKIDRQG-----AVAVASTASGAPTTASAPASSASAGPNOCVICYL-----RVLSC 638
1328 KGVDMNPDELKHILLACASASDKKRTYPPKKNPVPLKQTVQPKNGVVLNDSGKNAPRRMQ 1387
639 PRALRLHYGOGGERPFCKVC-----GRAFSTRGNLRAHFVGHKTSFA 682
1388 PKRLSFNV-ELGKMPNKLKLSLAKKKNLQVOKAILQKNRAKQKADLR-----DTSBA 1440
683 ARKONSPICKKRTTNAVTLQOHVHRHNLGQIPNG-----SALSEGGAOENSSQSPA 738
1441 --SSHICPYCDREFTYIGLSLKNHAAPSCPKPLSPSKRVYSHSSKGGHASSSSSDRNS 1498
739 SGP 741
1499 CHP 1501

RESULT 7
US-09-586-472-2
Sequence 2, Application US/09586472
Patent No. 632335

GENERAL INFORMATION:

APPLICANT: Huang, Shi
TITLE OF INVENTION: Retinoblastoma Protein - Interacting Zinc Finger Proteins

NUMBER OF SEQUENCES: 106
CORRESPONDENCE ADDRESS:

ADDRESSEE: Campbell & Flores LLP
STREET: 4370 La Jolla Village Drive, Suite 700
CITY: San Diego
STATE: California
COUNTRY: USA
ZIP: 92122

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/586,472
FILING DATE: 01-Jun-2000

CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 09/528,706
FILING DATE: 17-MAR-2000
APPLICATION NUMBER: US 08/516,859
FILING DATE: 18-AUG-1995
APPLICATION NUMBER: US 08/399,411
FILING DATE: 06-MAR-1995
APPLICATION NUMBER: US 08/292,683

FILING DATE: 18-AUG-1994
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-LJ 4130
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1706 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-586-472-2

Query Match 4.7%; Score 249.5; DB 4; Length 1706;
Best Local Similarity 22.1%; Pred. No. 2e-08;
Matches 186; Conservative 69; Mismatches 281; Indels 307; Gaps 38;

14 PCGEPAERGGDASEHNPOYCAK-----CCAQ-----FSDPTEFLAHQNSCCGTD 58
851 PCD---SEGEFEKENHLAQPAKKKKPTTCMLQKVLNEYNGVSLPTE-----TTP 898
59 PYWVITIGQENPSN-SSASAPRPE-GHSRSQVMDTEHSNPPDSGSGPPDPPTWGPERRG 116
899 EV-----TRSPCKCKSPDTPDPELGDSSCSVPTAESPEVVGSPSPLOT-----A 946
117 ESSSGQFLVATGTAAGGGGLILASPKLIGATPLPESTPAAPPPTPP-----PP 167
947 SLSSGQ-----LPLLTPTPESSPPCPDPLTVATPP 979
168 PGVSGHLNIPILLELKVQLQOHQOMTEQICQVLLGLGTVGAPASPELPGT 227
980 PLLPTVPLSHP-----SSDASPGQCSP 1002
228 GAASSTKPLPLPFS-----PIKP-----AQTGKTTASSSSSSSGAEPPKQAF 272
1003 FSNTTAGSPPLILSPVSPSPPIPPVEPLMSAASGPPTLSSSSSSSSS-----1052
273 HLXHPGSGQHPFSVGGVGRSHKPTAPSPALP-----GSTQDLIAS-PHIAFGTTGLAAQ 328
1053 ---FPSSS-----CSTSPSPPLSAVSSVSSGDNLEASLPVTFK-----1091
329 CLGAARGLEAAAPGLIKPKNGSGELGYEVISLSLEKPGGRHKCRFCAYKGSALQIH 388
1092 -----QEBSESEG-LKPKKEAPRAGGOSVQ--ETFSKNFICNVCESPLSIDLTKH 1141
389 LRSHTGERPYKCNVCGNRFTRGNLKVH-FHRH-----REKYPHV-QMNPHPVP 435
1142 LSVHAEMWPKCECVQLFKVKTDLSEHRFLHGVGNIFVCYCKKEFAFLCNLQOHORD 1201
436 EHLIDYVITSSGLPYGMSP-----PEKAEAEAGTPGGVVERKPLVASTTALATSESLLL 490
1202 LHPDEVCTHHEFESGTLRPONTDPSKANVEH-----1233
491 STGTSTAVAPGLPTFNKFLVLMKAVEP-KSKADENTPPGSEGSALAGVADSGSATRMOLSK 549
1234 -----MSLPE-----EPLETSREELNDSS-----ELYTTIKIMASG 1267
550 LVTSLSPLSMAL-LTNHLKSTGSPFPFY-VLEPLGASPSSETS-----KLQOLV 593
1268 IKTKDDPVRILGNQHYPSFKPPFQYHHRNPMGIGVATNFTTHNIDQFTTALICTKCG 1327
594 EKIDRQG-----AVAVASTASGAPTTASAPASSASAGPNOCVICYL-----RVLSC 638
1328 KGVDMNPDELKHILLACASASDKKRTYPPKKNPVPLKQTVQPKNGVVLNDSGKNAPRRMQ 1387
639 PRALRLHYGOGGERPFCKVC-----GRAFSTRGNLRAHFVGHKTSFA 682
1388 PKRLSFNV-ELGKMPNKLKLSLAKKKNLQVOKAILQKNRAKQKADLR-----DTSBA 1440

OY 683 ARQNSCPIQCKFTNAVTLOOHVHMLGGQIPNGC-----SALSEGGAOENSSQOSTA 738
Db 1441 --SSHICPYCDREFTYIGSLNKAHAFSCPKKPLSPSKRKRVSHSSKKGGHASSSSSDRNSS 1498
OY 739 SGP 741
Db 1499 CHP 1501

RESULT 8

US-09-528-706-2
Sequence 2, Application US/09528706
Patent No. 6468985
GENERAL INFORMATION:
APPLICANT: Huang, Shi
TITLE OF INVENTION: Retinoblastoma Protein - Interacting
TITLE OF INVENTION: Zinc Finger Proteins
NUMBER OF SEQUENCES: 106
CORRESPONDENCE ADDRESS:
ADDRESSEE: Campbell & Flores LLP
STREET: 4370 La Jolla Village Drive, Suite 700
CITY: San Diego
STATE: California
COUNTRY: USA
ZIP: 92122
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/528,706
CLASSIFICATION:
FILING DATE:
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 08/516,859
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/292,683
FILING DATE: 18-AUG-1994
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-LJ 1776
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ. ID NO. 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1706 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-528-706-2

Query Match 4.7%; Score 249.5; DB 4; Length 1706;
Best Local Similarity 22.1%; Pred. No. 2e-08;
Matches 186; Conservative 69; Mismatches 281; Indels 307; Gaps 38;

OY 14 PCGEFAERGDGDAEENHPQCAK-----CCAQ-----FSDPTFLAHONSCCTDP 58
Db 851 PCD-----SEKEFEKNHIAQAAKAKKPTTCLQVLLNEYNGVSLPTE-----TTP 898
OY 59 PYMVLIGGOENFSN-SSASSAPPE-GHSRSQVMDTEHSNPPDSSSGPPDPTWGPERRG 116
Db 899 EV-----TRSPSPCKSPDQPBDELGPDDSCSVPTASPEVVGSPSPLOT-----A 946
OY 117 EESSGCLVAATGTAAAGGGGLILASPKLGATPLPEESTPAPPPPPP-----PPP 167
Db 947 SLSSGG-----LPRLTPTPESSPPPCPPVLTVAATPPP 979
OY 168 PGVGGHNLNPLLEELRVLQQRQIHQMOMTEQICROVLLGLSGTQVAPASPSELPGT 227

Db 980 PLLPTVPLSHP-----SSDASPQCCPSP 1002
OY 228 GAASSTPLPLPFS-----PIKP-----AQTGKTTASSSSSSSSGAEPPKQAF 272
Db 1003 FSNHTAOSPPLPLSPVSPSPSPPIPPVEPLMSAASGPPTLSSSSSSSSSS----- 1052
OY 273 HLVPILGSQHPEFSVGVGRSHKPTAPAPALP-----GSTDOLIAS-PLIAPPGTGLLAQ 328
Db 1053 ---FPSSS-----CSSTSPSPPLSAVSSVSSGDNLEASLPAVTK----- 1091
OY 329 CLGAARGLAASAPGLKPKNGSGELGYGEVISTLEKPGGRHKCRCAKYFGSDSALQIH 388
Db 1092 -----QEESESEG-LKPKKEAPRAGGQSVQ--ETFSKNFICNCCESPFLSKDLTKH 1141
OY 389 LRSHTGERPYKCNVCGNFFTRGNLKVH-FHRH-----REKYPHV-QMNPHPVP 435
Db 1142 LSVHAEMPMFCEFCYOLFKVKTDLSEHFFLLHGVNIFVCSYCKKEFAFLCNQHQORD 1201
OY 436 EHLDYVITSSGLPYGMSVP-----PEKAEAGTPEGVERKFLVASTTALSNESLTL 490
Db 1202 LHFDEVCTHHEFESGTLRPQNFDPKNAVER----- 1233
OY 491 STGSTRVAPGLPTFNKFKVLMKAVEP-KSKADENTPPGSEGSIAIAGVADSGATRMQLSK 549
Db 1234 -----MPSLPE-----EPLTSREELNDSE-----ELYTTIKIMASG 1267
OY 550 LVTSLPSMAL-LTNHMLSTGSEPPPY-VLEPLGASSETS-----KIQOLV 593
Db 1268 IKTKDPVRLGNLQHYPSFKPPFQYHHRNPMGIGVATNFTTNIPQTFYAIRCTKCG 1327
OY 594 EKIDROG-----AVVASTASGAPTTAPAPSSASAPNQCIVCL-----RYLSC 638
Db 1328 KGVDMNELKHILLACASASASKKRTPKKNPVLKQOTVOPKNGCVVLDNSGKNAFRMGQ 1387
OY 639 PRALRLHYGGHGERPPCKKVC-----GRABSTRGNLRAHFVGHKTSPPA 682
Db 1388 PKRLSFNV-ELGKMSPKMLKLSALKKKNOVLQKAILQKRAAKQKDLR-----RYLSEA 1440
OY 683 ARQNSCPIQCKFTNAVTLOOHVHMLGGQIPNGC-----SALSEGGAOENSSQOSTA 738
Db 1441 --SSHICPYCDREFTYIGSLNKAHAFSCPKKPLSPSKRKRVSHSSKKGGHASSSSSDRNSS 1498
OY 739 SGP 741
Db 1499 CHP 1501

RESULT 9
US-08-459-568-4
Sequence 4, Application US/08459568
Patent No. 5811304

GENERAL INFORMATION:
APPLICANT: Huang, Shi
TITLE OF INVENTION: Retinoblastoma Protein - Interacting
TITLE OF INVENTION: Zinc Finger Proteins
NUMBER OF SEQUENCES: 93
CORRESPONDENCE ADDRESS:
ADDRESSEE: Campbell and Flores
STREET: 4370 La Jolla Village Drive, Suite 700
CITY: San Diego
STATE: California
COUNTRY: USA
ZIP: 92122
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/459,568
FILING DATE: 02-JUN-1995
CLASSIFICATION: 514
PRIOR APPLICATION DATA:


```

; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-399-411-4

Query Match      4.5%; Score 236; DB 2; Length 1719;
Best Local Similarity 18.1%; Pred. No. 1.6e-07;
Matches 230; Conservative 137; Mismatches 407; Indels 494; Gaps 50;

OY 27 EEHHPOVCAKCAQFSDPTE--FLAHQNSCC-----TDDPV----- 60
D 77 EYYVYRL-GMCMCIDATPEKGNMLRYVMACSGEONFLPLEINRAIYYTTLPIAGEE 135
OY 61 -WYIGQENP-----SNSASAPRPGHSRSQ-----VMDTE-HSNPP 98
D 136 LLYVMYGEDNPETIAAIEEERASARSKRSSPKSRKQKKKQENKNGNKIQDIDLKTPSEP 195
OY 99 DSGSSPPDPPTWPERGEESSGOFLVAATGTAAGGGGLIASPRLGATPLPPE-STPA 157
D 196 DFTSAMNRDSEAGKEDEEKPSASALEQPA-----TLQEVASQGEVPELATPA 243
OY 158 PPPPPPPPPGVSGHNLPLILEELRVLOQRQIHOMQTEQICQVLLGSLQTVGA 217
D 244 PAMEPQPEPERLEAACEVNDLGE-----EEEEDEDEEE----- 282
OY 218 PASPELPGTAASSTPKPLPLFSPRIKPAQTGKTTAASSSSSSSGAEPKQAFHLYHP 277
D 283 DDDDELEDEGEERASMP-----NENSVKEPE----- 309
OY 278 LGSQHFFSVGVGSHKPTPARSPALPSTDLIASPHLAFPTGTGLLAQCLGAAGLE 337
D 310 -----INDEKP-----EDLLEPKTTSEET-----LE 332
OY 338 --AASPGLPKPKNGSELGXEYISLEKPGRHKRCFAKVFSGDSALOIHLSH--T 393
D 333 DCESEVTPAMQIPR--TKEANGDVETFMF-----CQHCERKFTTKQGLERHMHITST 385
OY 394 GERPVYCNVCCNRETTKGNLKVHFRHR--EKYPHVQNNPHEPHELDVYITSSGLPYG 450
D 386 VNHAFKCKYCGKAFGTQINRRHRHRHAGLKRKPSQTLQP--SEDL-----ADGRASG 437
OY 451 MSVPRPKAEKEAGTGGGVEERKPLVASTTALSATESLTLLSTGSTRAVAP-----GL 502
D 438 ENV--ASKDSSPPSLGPDCLINMSERASODTINSSVVEENGVEKELHCKYCKVFGT 494
OY 503 PT-----FNKFLVMKAVERPSKADENTPPGSEGSATAGV----- 536
D 495 HTMMRRHQRRVHEHHLLPKGVRRKGGLEEROPRAQQAQTQNVYVSTEEDEEADVY 554
OY 537 -----ADSGATRMQLSKLVTSLPFWALLTNH 563
D 555 IMDISSNISENLNYIDGKIQTNNNTSNCDVIEMESASADLVGINCLLT--PTVETIQN 612
OY 564 LKSTGSFPPFYVL--EPLGASPS----- 585
D 613 IKTT-QVPTEDLPEKPELSTNSEKKRRRTASPPALPKIKAEETSDPMVPSCSLSPLSI 671
OY 586 -----TSKIQLOLVE-----KIDRQGAVAVASTAGAPTS 615
D 672 STTEAVSFHKEKSVYLSKQLQLOTOQKLPPRAGISATEIKAGLYCVASAPASMLPYTS 731
OY 616 APA PSSASAPNOCVICLRYLSCPRALRLHYGOHG-GERPFKCVKSGRAFSRGNL--R 671
D 732 SREKRRRTSSPP-----SSP-----QHSPPALRDFGKPSDGKAAWTDAGLSKK 773
OY 672 AHFVGHKTSFA-----ARAQNSCPIC-----OKKFT-----NATYLOOHVNRHLGGQI 714
D 774 SKLESHSDSPAMSLSGDERETVSPPCDEYKMSKEWTSASAFSVCNQDPL-----DL 827
OY 715 PNGSALSEGGAA-----QENSSQSTASG----- 740
D 828 SSGVKKQKAEKGTGKTPYQWESVLDLSVHKHKHCSDESEKKEKSHSVQPTCSAVVKKRKPPTC 887
OY 741 -----PGSFPOPOSOQPSPEEMSEBEED-----EEEEEDVTD 774
```

```

D 888 MLQKVLINENYNGIDLPEVNPADGTRSPSPCKSLDAQPDDPLGSGCFPAPTVESTPDVCP 947
OY 775 EDLSIAGSGSESGE---KAISVRGDSSEVSGAEAEVAT-----SVAAPT 816
D 948 SSPALQPTSSLSSGGLPPLLIPTDPSSPPCPVLTVATPPPLLPVLPAPSSASAPHP 1007
OY 817 VKEMDSNEKAPQHTLP-----PPPPP----- 838
D 1008 CPSPLSMATA-QSPPLPILSPVSPSPSPPIPVPEPLMSASAPGPTLSSSSSSSSSFS 1066
OY 839 ----DNLDHPQMEQGSTDVSGAMEEAKLEGISSPMALTOGC---EGTSPRLVEELN 890
D 1067 SSSSSSPSPPLSAISSVSSGDNLEASLPMISFKOELEENGLKPREEPQSAEQDVV 1126
OY 891 LPEAKMRDPGSSGKKAQCEVCGQSFPTOTALBEHOKTHPRDGLFTVCVFCRQGLDRATL 950
D 1127 VQETFNKN-----FYGVNCESPPLSTIKDLTKHLSIHEEWP-FKCEFCVQLFKDKTDL 1178
OY 951 KKHMLAH 958
D 1179 SEHFRFLH 1186
```

```

RESULT 11
US-08-516-859A-4
; Sequence 4, Application US/08516859A
; Patent No. 6069231
; GENERAL INFORMATION:
; APPLICANT: Huang, Shi
; TITLE OF INVENTION: Retinoblastoma Protein - Interacting
; TITLE OF INVENTION: Zinc Finger Proteins
; NUMBER OF SEQUENCES: 106
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell & Flores LLP
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/516,859A
; FILING DATE: 18-AUG-1995
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/399,411
; FILING DATE: 06-MAR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/292,683
; FILING DATE: 18-AUG-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 1776
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9901
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1719 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-516-859A-4

Query Match      4.5%; Score 236; DB 3; Length 1719;
Best Local Similarity 18.1%; Pred. No. 1.6e-07;
Matches 230; Conservative 137; Mismatches 407; Indels 494; Gaps 50;
```


27 EEHHPQYCAKCCAFSDPTE--FLAHQNSCC-----TDPV----- 60
177 EYVYPLN-GWMCIDATDEKGMWLRVYNNACSGEONLPLEINAIYKTLKPIARGEE 135
61 -WYIGGOENP-----SNSASAPREHRSQ-----VMDTE-HSNPP 98
136 LLYWYNEDNEIAAIBEERASARSKRSSPKSKKKKSOENKKNKIDQIQTLKTSER 195
99 DSSSSGPPDPWGPERRGESSGQFLVAATGTAAGGGGLLASKLCLATPLPPE-STPA 157
196 DFTSANMRDASBGPEDKEKPPASALEQPA-----TLQEVASQEVPPELATPA 243
158 PPPPPPPPPPPGSGSHNIPLEELRYLQORQIHOMOMTEQICQVLLLSLQTVGA 217
244 PAMPPQPPERDLRLAACEVNDLGE-----EEEEDEDEEE----- 282
218 PASPSELPGTGAASSTKPLPLPSPIKPAQTGKTATSSSSSSSSSGABPPKQAFHLYHP 277
283 DDDDELDEDEEEASMP-----NENSVKRPE----- 309
278 LGSHPSPVSGVGRSHKPPAPSPALPGSTDOLASPHLAPPGTTGLLAOCLGARGLE 337
310 -----IRCKDEK-----EDLLEPKTSEET-----LE 332
338 --AASPGLKPKNNGELGYGEVVISLEKPGGRHKCRFCAGKFGSDSALOHLRSH--T 393
333 DCEVNTAMQIPR--TKREANDVFETFMPP-----COHCKRKTTKOGLEHMHMHIIST 385
394 GERPKYCNVCNRRFTTGNLKVYFHRHR--EKYPHVOMNHPVPEHLDYVITSSGLPYG 450
386 VNHAFFKCKYCKAGFCTQINRRHRHRHEAGLKKRPSQTLGP-----SEDL-----ADGKASG 437
451 MSVPPEKAEAEAGTPGGVERKPRIVASTALSALESITLSTGTSTAVAP-----GL 502
438 ENV---ASKDSSPSPLDPCDIMNSEKASQDTINSSVVEENGVEKELHPCKYCKKVFGT 494
503 PT-----FNKFLKMAVEPKSKADENTPPGSEGSALIGV----- 536
495 HTNMRHRQRARHERLIRKGVARKGLLEFPRAEQAOATQNVYPSTEREEGEADVY 554
537 -----ADSGATRMQSLKVLTSPLSMALLNH 563
555 IMDISSISENLANYIDKIQTNNTNSNCDVIMESASADLYGINCLLT--PVTEITQN 612
564 LKSGSPFPFVYL--EPGASGSE----- 585
613 IKTT-QVPTMEDLPKEPLGSTNSEAKKRRTASPPALPKIKAEITDSDPVVPSGSLPLSI 671
586 -----TSKLOQLVE-----KIDROGAVAVASTASGAPTS 615
672 STEEAVSFHKKEKSVYLSKLLQLODKITPPAGISATETAKLQPCVSAASMLPVTS 731
616 AAPRSSASGENOCVYICRLVSLCPRALRLHYGQH--GERPKCYVCGRAFSTRGNL--R 671
732 SREKRRRTSSPP-----SSP-----OHSPALRDFGKPSDGRAAWTDAGLTSSK 773
672 AHVVGHTSPA-----ARAQNSCPTC-----QKFT-----NAVITQOHVBMHGGOI 714
774 SKIESHSDSPWMSLSGRDERETVSPCFDEYKMSKEWTASSAFSSVCNOOPL-----DL 827
715 PNGSALSEGGA-----OENSSEOSTASG----- 740
828 SSGVKQAKABEGKTPVQWESVLDLSYHKHCHSDEGEKFKESHVQPTCSAVKRRPTTC 887
741 -----PGSEFPOQSOQPSPEEMSEEEED-----EEEEVDYD 774
888 MLQKVLINEXYNGIDLPVENPADGTRSPCKSLAQPDPDLGPGSGFPAPYESTPVDOP 947
775 EDSLAGSGSEGGE--KAISVKGDSSEVSGAEEVAT-----SYAAPTT 816
948 SSPALQPTSLSSGQLPPLLIPTDSSPPCPQVLYATVTPPPLLPYVLPAPSSSASPSHP 1007

QY 817 VKEMDSNEKAPQHTLP-----PPPPP----- 838
Db 1008 CPGPLSNATA-QSPLPLISPIVSPSPSPIPVPEPLMSAASGPPPTLSSSSSSSSSSPS 1066
QY 839 ----DNIDHPQPMQGTSDVSGAMEEAEKLEGISSPMALITQEG----EGISTPLVEELN 890
Db 1067 SSSSSSSPSPLAISVSVSSGDNLEASLPMISFQOELENEGLKPREEQSAEAOVY 1126
QY 891 LPEAMKDPGESSGRKACVCGQSPFYOTALBEHQTHPKDGPFTGVCFGQGFDRATL 950
Db 1127 VQETFNKN-----FVNCVCSPLSTKDLTKHLSTHAEFWP-FKCFECVQLFKDKTDL 1178
QY 951 KKHMLLAH 958
Db 1179 SEHRFLH 1186

RESULT 12
US-09-586-472-4
; Sequence 4, Application us/09586472
; Patent No. 632335
; GENERAL INFORMATION:
; APPLICANT: Huang, Shi
; TITLE OF INVENTION: Retinoblastoma Protein - Interacting
; Zinc Finger Proteins
; NUMBER OF SEQUENCES: 106
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell & Flores LLP
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/586,472
; FILING DATE: 01-Jun-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 09/528,706
; FILING DATE: 17-MAR-2000
; APPLICATION NUMBER: US 08/516,859
; FILING DATE: 18-AUG-1995
; APPLICATION NUMBER: US 08/399,411
; FILING DATE: 06-MAR-1995
; APPLICATION NUMBER: US 08/292,683
; FILING DATE: 18-AUG-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 4130
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1719 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-09-586-472-4

Query Match 4.5%; Score 236; DB 4; Length 1719;
Best Local Similarity 18.1%; Pred. No. 1.6e-07;
Matches 230; Conservative 137; Mismatches 407; Indels 494; Gaps 50;
QY 27 EEHHPQYCAKCCAFSDPTE--FLAHQNSCC-----TDPV----- 60


```

77 EYVYVNL-GWMCIDATDPEKGNMLRYVWMAACSGEQNLPLEINRAIYYKTLKPIAPGEE 135
YY -WYIIGGQENP-----SNSASAPRPRGHSRSQ-----VMDTE-HSNPP 98
136 L1VWVNGEDNPETAAIAIEERASARSRSSPKSRKGKKSOENKNGKNIQDQLKTSSEP 195
99 DSGSSGPPDPTWGPERRGESSGOF1VAATGTAAGGGGLILASPLATPLPE-STPA 157
136 DFTSAMRDSABEGPKDEEKRKSALDOP-----TLOVASGEVPELATPA 243
158 PPPPPPPPPVGGSHLNP1LEELRVLDQRO1HOMQTEQICROVLLGLSGTVGA 217
244 PAMEPQPEPDERLEAACEVNDLGE-----EEEEEEDEEE----- 282
218 PASPSLPCTGAASSTPLPLFLFSPKPAQTKTASSSSSSSSSGAEPKQAFHLIYHP 277
283 DDDDELEDEGEERASMP-----NENSVEPE----- 309
278 LGSQHPEFVGVRSHKPTPARSPALPGSTDLIASPHLAFPTGTLLAQLGARGLE 337
310 -----IRCDER-----EDLEPKTTSEET-----LE 332
338 --AASPGLKPKNGSELGYGYV1SSLEKPGRHKRCFAKVFSGDSALOHLRSH--T 393
333 DCEVVPRAQIPR--TKEANGDVETFMFP-----CQHCERKFTTQGLEHMH1HIST 385
394 GERPYKCNVCNRTTGNLKVHFRHR--EKYHVQMPHVPREHLDYVITSSGLPYG 450
386 VNHAFKCKYCGKAGTQ1NRRRHRHREHAGLKRKPSQTLQ--SEDL-----ADGRASG 437
451 MSVPKEAKEEACGPGGVERKPLVASTALATESLTLTSTSTAVAP-----GL 502
438 ENV-----ASKDSSPPSLGPDCLIMNSEKASQDTINSSVVEENGEVKELHCKYKAYGT 494
503 PT-----FNKVLMAKAVEPKSKADENTPPGSEGSALAGY----- 536
495 HTNMRHRQRYVHERHL1PKGYARKGGLEROPPAFOAONVYVPSTPEEGEADVV 554
537 -----ADSGATRMOLSKVLTSLPSMALTNH 563
555 IMDISSISENLNY1DCK1QTNNTSNCDV1EMESASADLYGINCLLT--PVYETIYON 612
564 LKSTGSPFPYVL--EP1GASBPSE----- 585
613 IKTT-QYRVPVTEDLKREP1GISTNSEAKKRTTASPPALPKIAENDSDMVPSCSLPLSI 671
586 -----SKLOOLVE-----KIDROGAVAVASTAGAPPTS 615
672 STTEAVSFHKKEKSVYLSK1KOLLQ1TODKLT1PPAGISATE1AKLGPYCVSAPASMLPVTS 731
616 APAPSSASGNCV1L1RVLSCPRALRLHYGHG--GERPFKCKVCGRAFTRGNL--R 671
732 SRPKRRKTSPP-----SSP-----OHSPALRDFGKPSDGAATDAGLTGSK 773
672 AHFVGHKTS1PA-----ARAONSCPTC-----QKFT-----NAVTLQOCHVRMLGQI 714
774 SKLESHSDSPAMSLSGRDERETVSPRCFDEYKMSKEWTASSARSVCNOQPL-----DL 827
715 PNGGSALSEGGGA-----OENSSEQSTASG----- 740
828 SSGVKKAKKEGKTKTVOMESVLD1SVHKHKHCSDEKKEFKESHVQPTCSAVKKRPTTC 887
741 -----PSTPPOSOQOPSPREEEMSEEEED-----EEEDVTD 774
888 MLQKVL1LNEYNGIDL1PVENPADGT1RSPSPCKSL1EADPD1GPGSGFPAPV1ESTPDVCP 947
775 EDSLAGRSSESGE---KA1SVRGDSEEVSGABEEVAT-----SVAAPT 816
948 SSPALQPTSLSSGGLP1LL1P1DPSPPCPPV1TVATPPPL1PTV1LPA1SSASAPAP 1007
817 VKEMDSNEKAPDHTLP-----PPPPP----- 838
1008 CSPPLSNMTA-QSP1L1LSP1TVSPSPSP1P1PVP1LMSA1SPGPT1LSSSSSSSSSSSFS 1066

```

```

QY 839 -----DNLDPQPMEGTSDVSGAMEEAKLEGISSPMALTOEG-----EGSTPLVEELN 890
DB 1067 SSSSSSSPPPL1A1SSVSSGDNLEASLPM1STFOE1LENGLKPREPQSAEDDV 1126
QY 891 LPEAMKRDPGESSGRKACEVCGQSPPTOTAL1EHOKTHPKDGLPTCVFCRQGLDRTL 950
DB 1127 VQETFNKN-----FVGNVCSPLSTKDLTKHLS1HA1EWP-FKCEFCVQLFKDKTDL 1178
QY 951 KKHMLAH 958
DB 1179 SEHREFLH 1186

RESULT 13
US-09-528-706-4
; Sequence 4, Application US/09528706
; Patent No. 6468985
; GENERAL INFORMATION:
; APPLICANT: Huang, Shi
; TITLE OF INVENTION: Retinoblastoma Protein - Interacting
; NUMBER OF SEQUENCES: 106
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell & Flores LLP
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patentln Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/528,706
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/516,859
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/292,683
; FILING DATE: 18-AUG-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 1776
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1719 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-528-706-4

Query Match 4.5%; Score 236; DB 4; Length 1719;
Best Local Similarity 18.1%; Pred. No.1.6e-07;
Matches 230; Conservative 137; Mismatches 407; Incls 494; Gaps 50;

```



```
196 DFTSANMRDASBPCKDEEKPASALQPA-----TLOEVASQEVPELATPA 243
158 PPPPPPPPPGVSGHLNIPILILELRVLQOROIHQOMTPOICRQVLLGSLGQTVGA 217
244 PAMEQPEPDERLEAANAACEVNDLGEF-----EEEEEEDEEE----- 282
218 PASPELPTGAASSTKPLPLPFSPIKPAOTGKTASSSSSSSSSGAEPKQAFHLYHP 277
283 DDDDELEDEGEAEAMF-----NENSVKRE----- 309
278 LGSQHPFVSGVGRSHKPPRPPSPALPGSDQLASPHLAPGTGGLAACLGAARGLE 337
310 -----IRDEKP-----EDLLEPKTISEET-----LE 332
338 --AASPGLLKPNKSGELGYEVISLSEKPGGRHKRCFAKFGSDALQIHLRSH--T 393
333 DCEVTPAMQIDR--TKEANGVFETFMF-----CONCERKFTTKGLEHMHHT 385
394 GERPYKAVCGNRFTTRGULKYVFNHR--EKYPHOMNHPVPEHLDYVITSSGLPYG 450
386 VNAHFKCYCGAFGQJINRRRHERHREAGLKRKPSOTLP--SEDL-----ADGKASG 437
451 MSVPPEKAEERAGTGGVERKPLVASTALSATESLTLSTGTSTAVAP-----GL 502
438 ENV--ASKDDSPPSLGPDCILMSEKASODTIINSSVEENGEYKELHPCKYCKKVFGT 494
503 PT-----FNKFLMKAVPEKSKADENTPPGSEGSATAGV----- 536
495 HTNMRHORVHERHLIPKGVRRKGLLEPQPAEQAOATQNVYVSTPEPEEGADVY 554
537 -----ADSGATRMQLSKVLTSPLSMALLTNH 563
555 IMDISSNISEMLNYIDGKIOTNNNTNCDVITEMESADLYGINCLLT--PVYVEITON 612
564 LKSTGSPPEPYVL--EPLGASPE----- 585
613 IKTT-QYVPTEDLPREPISGTNSEAKKRRTPASPPALPKIKAEIENDSPVBCSLSLPIST 671
586 -----TSKLOOLV-----KIDROGAVAVASTAGAPTTS 615
672 STEEAVSFHKESVYLSSKLOLLOTODKLTPAGISATEIARLGPVCSAPASMLPYTS 731
616 APAPSSASGPNOCVIGLRVLSCPRALRLHYQH--GERPFKCVCGRAFSTRGNL--R 671
732 SRKRRRTSSP-----SSP-----QHSPLARFGKPSDCAAWTDAGLTSKK 773
672 AHFVGHKTPA-----ARAONSCPIC-----OKKFT-----NAVTLQOHVRLHLSGOI 714
774 SKLESHSDSPAMSLSGRDERETVSPCFDEYKMSKEMWTASSAFSSVONQPL-----DL 827
715 PNGSALSEGGA-----QENSSEOSTASG----- 740
828 SSGVOKAKAEGTGKTPVQWESVLDLSVHKHCSDSEKKEFKESHVSVPSCSAVKRRKPTTC 887
741 -----PGSFPOQOSPPEEEMSEBEED-----EEEEPYTD 774
888 MLOKVLINEXNGIDLVPENPADGTRSPSPCKSLEAOPDDPLGSGCPAPVVESTPDVCP 947
775 EDSLARGSGSGE--KAISVRGDSSEVSGAELEVAT-----SVAAPT 816
948 SSPALOTPSLSSGOLPRLPTDSSPPCPVPLVATPPRLPTVPLPAPSSASAPHR 1007
817 VKEMDSNEKAPHTLP-----PPPPP----- 838
1008 CSPPLSNATA-QSPLPLISPTVSPSPSPPIPVPEPLMSAASPPTLSSSSSSSSSSSFS 1066
839 ----DNLDHPOPEOQTSDVSGAMEEFAKLEGISSPMALTOEG-----EGSTMPLVELN 890
1067 SSSSSSSSPSPPLISALISSVYSSGDNLEASLPMISFKQELEHGLKPRREPQGAEDDV 1126
891 LPBAKKDGESSGRKACEVCGSGFPYOTALEHOKTHPKDQPLFTCVCRQGLPRATL 950
```

```
Db 1127 VOETFNKN-----FVCNVCSEPSFLSKIDLTKHLSTHAEMP-FKCEFCVOLFKDKTDL 1178
QY 951 KKHMLLAH 958
Db 1179 SEHRLFLH 1186
```

RESULT 14

```
US-09-262-773-4
; Sequence 4, Application US/09262773
; Patent No. 6225451
; GENERAL INFORMATION:
; APPLICANT: Ballinger, Dennis G.
; APPLICANT: Ding, Wei
; APPLICANT: Wagner, Susanne
; APPLICANT: Hess, Mark A.
; TITLE OF INVENTION: CHROMOSOME 11-LINKED CORONARY HEART DISEASE
; FILE REFERENCE: Myriad 3
; CURRENT APPLICATION NUMBER: US/09/262,773
; NUMBER OF SEQ ID NOS: 210
; SOFTWARE: Patentlin Ver. 2.0
; SEQ ID NO 4
; LENGTH: 640
; TYPE: PRT
; ORGANISM: human
US-09-262-773-4
```

```
Query Match 4.4%; Score 234.5; DB 4; Length 640;
Best Local Similarity 20.2%; Pred. No. 5,7e-08;
Matches 129; Conservative 55; Mismatches 233; Indels 223; Gaps 22;
```

```
QY 145 LGAPPLPE--STAPPPPPPPPPGVSGHLNIPILILELRVLQOROIHQOMTPOIC 202
Db 144 LGAEPSENELODPVQSTPEOSPEETQSPDLGAP--AEQRHQBELQTLDESE-- 197
QY 203 ROVLLGSLGQTVGAPASPELPGTGAASSTKPLPLFSPIKPAQOTGKTASSSSSSSS 262
Db 198 -----VPYBEDP-DLPAERSGDSSENVALLTALSQGLVTRDVAVCSQDDMS 244
QY 263 GAEPKQAFHLY-----HPLGSHQPFVSGVGRSHKPPRPPSPALPGSTDO 309
Db 245 DLDPQKFEYGEYVLEDDGIVLSLFP1--PRPDEISQV-REBEPV-VPDIOEQFOE 300
QY 310 LIASPHL--AFPGTGLLAQCLGAA-RGLEAASGLKLP-----KNGSGE 353
Db 301 ----PELISFTYTGGRSKDEDECELEDLSEDIHNPVLEPELHQTPDWEIVFEDNPGR 356
QY 354 LG--YGEVVISLEK-----PGRRHKRCFAKVFSGDSALQIHLRSHHTGERP 397
Db 357 LNERFGTINISQVNSFVNLRETTPVHPLTGRHNDGVCGRKSPFONSHLVRLRTHHTGERK 416
QY 398 YKCVGCGRFTTRGULKYVFNHRREKYPHVOANPHVPEHLDYVITSSGLPYGSAVPEK 457
Db 417 YKCEGKSTYRSSHLARHOKVKNMADY-----KYLNR 451
QY 458 ABEAGTPEGGVERRKPLVASTALATESLTLSTSTAVAPGLPTFNKFLMKAVEPK 517
Db 452 KMLEETSPVTOAERTP-----SVEKP 472
QY 518 SKADENTPPGSEGSALAGVADSGATRMQLSKLVYSLPSMA--LTLNHLKSTGSFPP-- 572
Db 473 YRCD-----DCGKHFR-----WTSDLVNRHQRTNKEKPEPCT 504
QY 573 -----PVYLEPLGASPESTSKLOOLVEKIDROGAVAVASTAS 609
Db 505 ICKSFSQKSVLTHQRIHLGKRPVLCGEGDEFSHRY----- 544
QY 610 GAPTTSAPAPSSASGPNOCVIGLRVLSCPRALRLHYGONGGERPPCKVCYGRAFTSGN 669
Db 545 ----LAHRTTHAAEELIYLCSECGRCFTTHSAFAKHLKRHASVRCPCNCEGKSFSFRD- 598
```


467 KCRYCDVAFVHERYALI-----QHOKSHKNEKRFKCDQCYACHQERHMV 511
y 644 LHYGOGGERPFCKYCGRAFTSTRGNLRAHFVGHKTSPPAARQNSCPICQKFTTNAVTLQ 703
y 512 MHRKTHTEKPYACSHCDKTFRKQOLDMHFKRYHDPNFVPAAFVCSKCGKFTTRNTMA 571
y 704 QVHRMHLGQIIPNGSALSGGAOENSSQSTASPGSFPOQSPPEEEMSEEE 763
y 572 RHA-----DNCGLDGGEG---ENGGE--TKKGKRG---RRKMRSKEDSSDSEE 614
y 764 EDEEEDVTDDESLAGRSGESGGEKALSVRGDSEEVSGAEAEVATVVAAPTVKEMDSN 823
y 615 NAEPLDDNDEDE-----ETAVEIE-----AEPEVGAEPAPPPSKRRGR 655
y 824 EKAPQHTLPPPPPPNDLHPQPM-----EGTSDVSGAMEEAKLEGISSPMAALTQ 877
b 656 -----PPGKAATQTKOSQPAIIQVEDQNTGEIENIIVEYK-----EPDAETVEE 701
y 878 GEGSTPLVEELN 890
b 702 EEAPAVEAPN 714
RESULT 17
CT-US95-08429-5
Sequence 5, Application PC/TUS9508429
GENERAL INFORMATION:
APPLICANT:
TITLE OF INVENTION: CTCF
NUMBER OF SEQUENCES: 21
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/08429
FILING DATE: 15-JUN-1995
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/261,680
FILING DATE: 17-JUN-1994
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Parmelee, Steven W.
REGISTRATION NUMBER: 31,990
REFERENCE/DOCKET NUMBER: 14538A-11-1PC
TELECOMMUNICATION INFORMATION:
TELEPHONE: 206-467-9600
TELEFAX: 415-543-5043
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 728 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
CT-US95-08429-5
Query Match 4.3%; Score 228.5; DB 5; Length 728;
Best Local Similarity 21.5%; Pred. No. 1.7e-07;
Matches 119; Conservative 58; Mismatches 221; Indels 155; Gaps 23;
y 364 EKPGRHRCRPFCAKVGSDALQIHLRSHTEGRRPKVCNVCGRFTTRGNLKVHHRREK 423
y 291 ERP---HCHLGRARRTVLLRNHLNLTHTGTRPKCPDCDMAFVTSGEIV---RHR-R 342
y 424 YPHVQMNHPVPEHLDYVLTSSGLPYGMSVPEPEKAEAGTGGGVERKPLVASTATLSA 483
y 343 YKHHEKPFKC-SMCDYA-----SVEVSKLKRHRSTG---ERPFQCSLCSYAS 388
y 484 TESLTLSTGTSTAVAPGLETFNKFLVLMKAVEPRKSKADENTPPGSEGSAT---AGVADS 539
y 389 RDTY-----KLKRMHRTSHGKPYEYGLCHARTOS 419

QY 540 GSATRMOLSKLTVSLSPWALLTNHLKSTGSPFPYV-----LEPLGA-----SPSETS 587
Db 420 GT---MKMH-----LQKHTENVAKFHCPCDFTVIARKSDGLVHLRKQHSYIEQG 466
QY 588 KLOQLVEKIDRGAVAVASTASGAPTTSAFAPSSASGPNQ---CVICLRVLSCPALR 643
Db 467 KCRYCDVAFVHERYALI-----QHOKSHKNEKRFKCDQCYACHQERHMV 511
QY 644 LHYGOGGERPFCKYCGRAFTSTRGNLRAHFVGHKTSPPAARQNSCPICQKFTTNAVTLQ 703
Db 512 MHRKTHTEKPYACSHCDKTFRKQOLDMHFKRYHDPNFVPAAFVCSKCGKFTTRNTMA 571
QY 704 QVHRMHLGQIIPNGSALSGGAOENSSQSTASPGSFPOQSPPEEEMSEEE 763
Db 572 RHA-----DNCGLDGGEG---ENGGE--TKKGKRG---RRKMRSKEDSSDSEE 614
QY 764 EDEEEDVTDDESLAGRSGESGGEKALSVRGDSEEVSGAEAEVATVVAAPTVKEMDSN 823
Db 615 NAEPLDDNDEDE-----ETAVEIE-----AEPEVGAEPAPPPSKRRGR 655
QY 824 EKAPQHTLPPPPPPNDLHPQPM-----EGTSDVSGAMEEAKLEGISSPMAALTQ 877
Db 656 -----PPGKAATQTKOSQPAIIQVEDQNTGEIENIIVEYK-----EPDAETVEE 701
QY 878 GEGSTPLVEELN 890
Db 702 EEAPAVEAPN 714

RESULT 18
US-08-040-548-1
; Sequence 1, Application US/08040548
; Patent No. 5763209
; GENERAL INFORMATION:
; APPLICANT: Sukhame, Vikas P.
; TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO THE
; TITLE OF INVENTION: FUNCTIONAL DOMAINS OF DNA BINDING PROTEINS
; NUMBER OF SEQUENCES: 67
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: 321 No. 5763209th Clark Street, Suite 800
; CITY: Chicago
; STATE: Illinois
; COUNTRY: U.S.A.
; ZIP: 60610
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/040,548
; FILING DATE:
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Coughlin, Daniel F.
; REGISTRATION NUMBER: 36,111
; REFERENCE/DOCKET NUMBER: arcd067
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (312) 744-0090
; TELEFAX: (312) 245-4961
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 533 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-040-548-1
Query Match 4.3%; Score 226; DB 1; Length 533;
Best Local Similarity 22.1%; Pred. No. 1.7e-07;


```

REGISTRATION NUMBER: 36,111
REFERENCE/DOCKET NUMBER: arc0067
TELECOMMUNICATION INFORMATION:
TELEPHONE: (312) 744-0090
TELEFAX: (312) 245-4961
INFORMATION FOR SEQ. ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 533 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-466-344-1

Query Match      4.3%; Score 226; DB 1; Length 533;
Best Local Similarity 22.1%; Pred. No. 1,7e-07;
Matches 124; Conservative 66; Mismatches 190; Indels 180; Gaps 26;

QY 70 PSNSSASSAPRPGHSHSQVMDTEHSNPPDSSGSGPDPDTWCGERGESSGQFLVAATG 129
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 24 PHSPTMDNYPKLE-----EMLLNSGAPOFLCAAGTP-----EGSGGSSSTS 67
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 130 TAAAGGGGGLLAPKLGATPLPESTAPRPPPPPPPPPPGVSCHLNIPLLELRVLCQ 189
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 68 SGGGGGGG-----SNSGSAFNPQGEPEQP-----YELLTE-SFSIALNNE 110
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 190 ROIHQMQTEQICROYLLGSLQTVGAPASPELPGTGAASSTKPLPLPSPF-----KP 245
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 111 KAMVEISYSQTR-----LPIYTGRLSLERA--PNSGNTLMPPELFLSLVSGLSVMNP 164
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 246 AQTGKTTASSSSSSSSGAEPKQAFPHLYHPLGSOHPFSVGVGVSQSHKPT-PAP----- 299
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 165 PHSSSAPSPSAASSSSASQSP-----PLSCAVPNSDSSPIYSADPTFPPTNDIF 215
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 300 -----SPALCGSTQOLLSPHLAPGTTG-----LLAQ-----CLGAA-----RGL 336
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 216 PEPQSAFPEFSAGTALOYPPRAYPATKGGFOYPMIPDYLFPPQOGDLSLCTPDQKPOGL 275
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 337 EAAASGGLKLP-----KNGSGELGYGEVI--SLEKPG----- 367
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 276 ENRTQGPSLPLPLTIKAFATQSGSDLKALNNTYOSQLKPSMKRKYPNRPSKTPHERP 335
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 368 -----GRH-----KCRFCAYFGSDSALQIHLSHTGERPYK 399
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 336 YACPVESCRRFRSRDELTRIIRIHTGQKPFQCRICMRNFRSDHLTTHIRTHGEKPEFA 395
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 400 CAVCGKRF-----TRGLKLVHFHHRER-----YHVNQNNHPVPEILD 439
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 396 CDICGKRFARSDERKRRTKTHLRQDKRADKSVASPAASSLSSYPSPVATSYSPATTS 455
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 440 Y---VTSSGLPLGMSVPEPEKAEAGTGGGVGERKPLVAST-----TALSATESL 487
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 456 FESPVPTTSSSP-GSSLYPSPA--HSGFP-----SPSVAITTPASVPAPPEYOVSSPESA 506
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 488 TLLST-GTSTAVALGDLPTFN 506
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 507 GVSSSFSTSTGLSDMTATFS 526
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 20
5206152-2
Patent No. 5206152
Applicant: SUKHATME, VIKAS P.
Title of Invention: CLONING AND EXPRESSION OF EARLY GROWTH
REGULATORY PROTEIN GENES
NUMBER OF SEQUENCES: 18
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/249,584
FILING DATE: 26-SEP-1988
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 179,587
FILING DATE: 08-APR-1988
(SEQ ID NO:2;

```


LENGTH: 533

206152-2

4.3%: Score 226; DB 6; Length 533;

Query Match Best Local Similarity 22.18; Pred.No.1.7e-07; Matches 124; Conservative 66; Mismatches 190; Indels 180; Gaps 26;

```
70 PNSSASAPRPEGHSQVMDTEHNSPPDSSGSGPPDPWGERGEESSGFVLAATG 129
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
24 PHSPTMDNYKLE-----EMLLSNGAPQFLGAGTP-----EGSGGSSSTS 67
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
130 TAAAGGGGGLILASPLKATPLPESTPAPPPPPPPPPGVSGHLINILILEIRVLOQ 189
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
68 SGGGGGGG-----SNSSSAFNPOGESEOP-----YEHLTTE-SFSDIALNNE 110
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
190 ROIHOMOMFOICQOVLLGSLGQVVGAPASPSLPGTGAASRKLPLPLPSPI----KP 245
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
111 KANVESYSPQTR---LPIITYGRPSLEPA--PNSGNTLMPEPLFSLVSGVSMWNP 164
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
246 AOTGKTATSSSSSSSGAEPKQAFPHLYHPLGSOHPFSGVGRSHKPT-PAP----- 299
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
165 PHSSSAPSPASSSSSASOSP-----PLSCAVPSNDSSPTYSAPTFPTPTDIF 215
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
300 ---SPALPGTDOLIASPHLAFPGTTG-----LLAQ---CLGAA---RGL 336
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
216 PEPQSOAFPSAGTALQYPPPAYPATKGGFOYPMIPDYLPQQGGDLGLHPDQKPFQGL 275
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
337 EAAASGLKLP-----KNGSELGGEVI--SLEKPG----- 367
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
276 ENRTOQPLPLTIKAPATQSGSODLALANTTYOSQLKPSRMKYPNRPDKTPPHERP 335
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
368 -----GRH-----KRCFCAYGSDSALQILHRSHTGERPYK 399
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
336 YACPVESCDRFSDELFRIRIHTGOKPPOCRICMRNRSRSHLTHITHTGEKRPFA 395
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
400 CAVCGNRF---TRGNLKVHFHHRK-----YPHVQMPHPVPEHL 439
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
396 CQICGRKFAASDERKRRHTKILRQKDKADKSVASPAASLSYSPVATSPAPATTS 455
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
440 Y---VTSSSLPGMSVPEKAEAGTGGVEKRPVAST-----TALSATESL 487
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
456 FSPSPVTSYSSP-GSSYTPSPA--HSGRP-----SPVATTPASVPAPFPQVSSFP 506
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
488 TILST-GTSTAVALGLETFEN 506
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
507 GVSSSFSTISGLSMTATFS 526
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 21
US-08-061-376-5
Sequence 5, Application US/08061376
Patent No. 6175000

GENERAL INFORMATION:
APPLICANT: Evans, Glen A.
APPLICANT: Djabali, Malek
APPLICANT: Selleri, Licia
APPLICANT: Parry, Pauline
TITLE OF INVENTION: CHARACTERIZATION OF A CHROMOSOME 11Q23
TITLE OF INVENTION: TRANSLOCATION BREAKPOINT ASSOCIATED WITH ACUTE LEUKEMIAS
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pretty, Schroeder, Brueggemann & Clark
STREET: 444 South Flower Street, Suite 2000
CITY: Los Angeles
STATE: California
COUNTRY: USA
ZIP: 90071

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
```

```
APPLICATION NUMBER: US/08/061,376
FILING DATE: 13-May-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Reiter, Stephen E.
REGISTRATION NUMBER: 31,192
REFERENCE/DOCKET NUMBER: P41 9387
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619)546-4737
TELEFAX: (619)546-9392
INFORMATION FOR SEQ. ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 3969 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-061-376-5
```

Query Match 4.3%: Score 226; DB 4; Length 3969;

Best Local Similarity 19.9%: Pred. No. 2.3e-06; Matches 245; Conservative 117; Mismatches 474; Indels 394; Gaps 52;

```
QY 2 AGETSSSRIGPCGEPAERGDASEEHHPQVCACCAQFSDPTEFLAHONSCCTD-PPV 60
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 452 APSGSSSEKSSAASQHSQMSDSSRSSPSVDTSQASERIQVLPERRSDPEVHP 511
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 61 MYIIGQENPSNSASS---APREGHSRQVMDTEHNSPPDSSGSGPPDPWGERRG 116
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 512 LPIQSPEENESNDRRSRYSVESFSGRTTKSLTLQSAPOOQTSSPPPLLTTPPPL 571
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 117 EESS-----GQFLVAATGTAAGGGGGLILASPKIGATPLPESTPAP 159
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 572 QPASTISDHTPMLPPIPLASPLPASTAPMOGRKS-ILKEPTRTWISL-KHSREPO 629
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 160 -----PPPPPPPPGVSG----- 173
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 630 YPSAKYAKEGLIRKPIFDNFRPPLPREDYGFASGFSASGTAAARLFSPLHSGTFRDM 689
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 174 HLNIPILERLYLOQRQIHOMQMTQICROYLLGSLQTYGAPRSPSLPGGAASS- 232
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 690 HRSPL-----LRAPRTPEAHSRIFEVTL-----PSRRTSAGTSSSG 729
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 233 ---TKPLPLESPIK---PAOTGKTATSSSSSSSGAEPKQAFPHL---YHPGGS- 280
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 730 VSNRRKRKRVFSPTRSPRBSHMRTRSGRLSSSELPLTPSSVSSLSISVPLATS 789
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 281 -----QHPPSYGVGV-----RSHKPTPAPSPALPGSTDOLIASPHLAFPGTGILA 326
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 790 ALNPFTFPSSHSLTQSGESAENQRKOTSAPAEFFSSS-----SPRPLFPWFTP--G 841
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 327 AQCGLAARGLAASAPLL----- 353
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 842 SQ---TERGRNKKAPBELSKDRADKSVKDKSHERDEREKREKREKRRKKS-- 896
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 354 LGYGEV-LSLEKPGGHKRCFCAYFGSDSALQILHRSHTGERPYKAVCGNRF--- 409
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 897 ----ETQSSALYPGVKSE---KYVGEDVATSSAKATGRKSSSHSGDITIVTL 949
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 410 -----RGNLKVHFHHRKREKYPHVQMPHPVPEHLIDYVITSSGLPYGMSVP 454
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 950 GDTTAVKTKILIKKGNL----- 968
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 455 PEKAEFAGTPGGVEKRPVASTALSATESLTLTSTGTAVALGCLPTFNKFV---LM 511
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 969 -EXTNLDIGPTAPSLERKTLCLSTPSSSTYVKSSTISGMSLAQADKLPTDTRVVASL 1027
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 512 KA-----VEPKSKADENTPPGSEGSALAGVADSGSATPMOLSKLTVSLPSMALLTHL 565
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1028 KAKAOLCKTEKSKSLQTDOPKAGOE---SDS-SETSVRGRIRIHVCRRAAVALGRKR 1082
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 566 STGSPFPVLEPLGASP-SETSKILOOLVERKIDROGAVAVAVASGAPPTTSAPAPSSAS 624
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
```


Db 1083 AV----FPDDMPTLSALPWEEREKILSSMGNDKSSIASSEDAEPLAPIKPIKPYTRNK 1138
OY 625 GPNOCYI-----CLRVLSCPRALHGHQGERPKVCVCGRAFSTRGLRAHFVGHKT 679
Db 1139 APOEPPVKKGRRRRCQCP-----GCQVEDCGVCTNCLD-----KKEFGGRNT 1183
OY 680 SPARAQNSCPICQ-----KKFTNAVTLQOHVRLHGLGOIPNGSALSSEGGAAQEN 731
Db 1184 KQCCMKRRCQNLQMMPSKAYLQKQAKAVKKKK-----KSKTSE-----KRD 1226
OY 732 SSEOSTASC---PGSFQPOPSQO--PSPPEEMSEEEDEEDVDTEDSLARGSESGG 787
Db 1227 SKESSVKKNVVDSSOKTTPSAREDPAPKSSSEPPRPKPVEEK--SEEGNVASAPRESQ 1284
OY 788 EKALSRGDSSEVSGAEAEVATSVAPPTVKEMDSEKARQHTLPPR-----PPRDLDH 843
Db 1285 ATTPARKSKKQVS--QPALVTPPOPTT---GPRKREVPKTPSPKKKOPPPES--G 1337
OY 844 POPMEGSTDVSGAMEBEAKLEGISSPMALTOEGEGSTPLVEELNLPKMKDPGE-- 901
Db 1338 PQSGKQKVAAPRSIPYKQPKKEKEKPPVKNQENAGTLNLTSLNSNGSSKQKIPADGV 1397
OY 902 -----SSGRKA-----CEVCGSFPPTQ 918
Db 1398 HRIIVDFKEDCEAENWEMGGLITSVPTPRVVCFLCASSGHVEFYVCQVCEEPF-HK 1456
OY 919 TALEEHQHTHPKDGPLET-----CVFCQGFDRATLTKKMLLHHQVPPFAHGPQNTA 973
Db 1457 FLELENER--PLEDOLENNCCRRCKFC-----HVCGRHO-----A 1490
OY 974 TLSLVP--GCCSSIPSPGLSP-FPRKDDPT 1000
Db 1491 TKOLLECNKCRNSYHRECLGPNYPTK--PT 1518
RESULT 22
US-08-475-844-9
Sequence 9, Application US/08475844
Patent No. 5972643
GENERAL INFORMATION:
APPLICANT: Lobanenkov, Victor V.
APPLICANT: Nelman, Paul E.
APPLICANT: Klenova, Elena M.
APPLICANT: Goodwin, Graham H.
APPLICANT: Filippova, Galina N.
APPLICANT: Collins, Steven J.
TITLE OF INVENTION: CTCF
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSER: Townsend and Townsend Kourile and Crew
STREET: One Market Plaza
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94105
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/475,844
FILING DATE: 07-JUN-1995
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/261,680
FILING DATE: 17-JUN-1994
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Parmelee, Steven W.
REGISTRATION NUMBER: 31,990
REFERENCE/DOCKET NUMBER: 14538A-11-1
TELECOMMUNICATION INFORMATION:

TELEPHONE: 206-467-9600
TELEFAX: 415-543-5043
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 727 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-475-844-9
Query Match 4.3%; Score 225.5; DB 2; Length 727;
Best Local Similarity 20.7%; Pred. No. 2.7e-07;
Matches 115; Conservative 61; Mismatches 207; Indels 173; Gaps 23;
OY 364 EKGGGRKRCPCAVFGSDSLQIHLSHTGEPYKCNVCNCFRTNGLKVHFRHREK 423
Db 291 ERP--HKCHLCGAFRTVTLRNHLNTHGTGRPHKCPDCDAMFVTSGLVY---RRR-R 342
OY 424 YPHVQMPHPVPEHLDYVITSGLPYGMSVPPEKAEAEAGTPEGGVKRPVASTTALSA 483
Db 343 YKHTHEKPFKC-SMCDVA-----SVEYSKLRHHSHTG---ERFQCSLCSTAS 388
OY 484 TESLTLTSTGTAAPGLPTFNKFLMKAVPEKSKADENTPPGSEGSAT---AGVADS 539
Db 389 RDTY-----KLKRHRTSHGEEKPYECYICHARTQS 419
OY 540 GSATRMQLSKLYTLSPWALLTNHLKSTGSPFPYV-----LEPTGA-----SPETS 587
Db 420 GT--MKMH-----ILQKHTENAKFHCPHCDTVIARKSDGLVHLRKQHSYIEQG 466
OY 588 KLOIVKIDROGAVAVASTASGAPTSAPAPSSASGPNQ---CVICLRLVSCPRLR 643
Db 467 KRCRYCAVHERKALT-----QHOKSHNEKFKCDQCYACQERHMT 511
OY 644 LHYGQHGGERPKCKVCGRAFSTRGNLRAHFVGHKTSAPARAQNSCPICQKFTNAVTLQ 703
Db 512 MHRTHTEGERPYACSHCDKTFROKOLDMHFKRYHDNPVPAFVCSKCGKTFRRTTMA 571
OY 704 QHVRMHGGQIPNGSALSSEGGAAQENSSBSTASGPSFPQPOQOOPSPPEEMSEEE 763
Db 572 RHADNCAG--PDG---VEG---ENGGETKSKRGRKKRKRKEDSSDSNAEPDL 619
OY 764 EDEEEDVDTEDESLAGSESGEKAISVRGDSSEVSGAEAEVATSVAPPTVKEMDSN 823
Db 620 DQWDEDE-----EPRAVEIPEPE----- 637
OY 824 EKAPQHTLPPPP-----PPDNIDHPQPMQGTSDV-----SGAMEE--EAKLEGI 867
Db 638 ---POVTPAPPAKRRGRPPGRTPQPK-QNOPATIIQVEDONTGAIENITVEVKKRPD 693
OY 868 SSPMALTOEGEGST 883
Db 694 AEPPAEGEEDQAQPAAT 709
RESULT 23
PCT-US95-08429-9
Sequence 9, Application PC/TUS9508429
GENERAL INFORMATION:
APPLICANT:
TITLE OF INVENTION: CTCF
NUMBER OF SEQUENCES: 21
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/08429
FILING DATE: 15-JUN-1995
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/261,680


```
QY      813 APTTVKEMDSNEKARQHTLPPPPPP-----PDNLDPQPMEOGTSVDSGAMBEAKLEGI 867
      452 SPVT-----TSYSPSAPATTSYSPVPTSFSSSGSSSTYSPSVH-----SGF 490
OY      868 SSPMAALIOEBGTSFPLVEELNLPEAMKKDPGESSGKRACEVCGSFPTOTALBEOHKT 927
      491 PSPSVAATYS-----SVPAPAF---PAQVSSFPSSSAVT--NSFSASTGLSDMTAT 534

RESULT 25
US-09-579-181-2
: Sequence 2, Application US/09579181
: Patent No. 6365372
: GENERAL INFORMATION:
: APPLICANT: Chirivia, John
: APPLICANT: Yaciuk, Peter
: TITLE OF INVENTION: SNE2 Related CBP Activator Protein (SRCAP)
: FILE REFERENCE: 16153-4247
: CURRENT APPLICATION NUMBER: US/09/579,181
: CURRENT FILING DATE: 2000-05-25
: PRIOR APPLICATION NUMBER: 60/136,620
: PRIOR FILING DATE: 1999-05-27
: NUMBER OF SEQ ID NOS: 17
: SOFTWARE: Patentl Ver. 2.0
: SEQ ID NO 2
: LENGTH: 2972
: TYPE: PRT
: ORGANISM: Human
: US-09-579-181-2

Query Match      4.2%; Score 224.5; DB 4; Length 2972;
Best Local Similarity 20.8%; Pred. No. 2e-06;
Matches 218; Conservative 78; Mismatches 317; Indels 435; Gaps 42;

OY      72 NSSASSAPRPEGHSRQVMDTEHSNPPDSGSGPPDPPTWGPERRGESSGQFLVAATGTA 131
      798 NMLQPVPRKQEGRTVVVV-----NNPRAPLGPVVRPPPGPELSAQPTGPPV----- 844
OY      132 AGGGGGLILASPRLGATPLPPESTPAPPPPP-----PPPPPGVSGHLNPLIL-EBLRVL 187
      845 ----POVLASLSLWVSASPAGPRLIPASRPBPVLLPILQPNNGS----LPQVLSPPLGVL 896
OY      188 QORQIHQMOMTEQICQOVLLLSLQGTVGAPASPSSELPGTGAASKRLLPLFSPPIKPAQ 247
      897 SGTSS-----RPPPTLSLKPFPAPVRLSPAPPPGSSSLKPL-----TYPPGY 940
OY      248 T-----GKTTASSSSSSSSSGAEPKQAFPHLYHPLGSHQHPFSVGVGGRSHKPTPASPAL 303
      941 TFPFAAATTTSTTAATATTAVP-----APTPAP----- 969
OY      304 PGSTDOLIASPHLAFPGTTGLLAAQCIGAARGLMAASPGLLKPKNGSGE-LGYGEVSS 362
      970 ----QRLLISPD-----MQARLPSEEVVSTIGQLASL 996
OY      363 LEKPGGRHKRCFCAKAVGSDSALOIHLSHTGERPYKCNVCNREPTTRG----NLKVHFH 418
      997 AQR-----VANAGSKPLTFQIQNKLLTLGAQVRLAVGQP 1034
OY      419 RHREKYPHYQMN-----PHVPEHLDVDYTTSSGCLPYGSVVP 455
      1035 RFLQAPPTVAVNNNTGVVYKIVRQAPRDGLTPVPPLAPAPR-----PSSGAPVAVNNPR 1087
OY      456 EKAEEBAGTPGGGVERRKPLVASTTALSTESTLTLSTGTSTVAPGLPFPNFKVFLKAVE 515
      1088 TLTTPGLPRTPLGTATAPM-----PTPLTVRP-----LTKLVH 1120
OY      516 PKSKADENTPGSE-----GSAIAGVADS-----GSATRMQLS--- 548
      1121 SPSPSVASASAPCAAPLTITSSPLHVPSLSLCPASSPMPIPNSSPLASPVSTSVPLSSSL 1180
OY      549 --KLVTSLPS-----WALLTNHLKSTG-----SFPFPVYLEELGAS 582
      1181 PISVPTTLPAAPASAPLTLTIPISAPLTVSASGAPLTLTSYMPPLAPVVAABGPPSLQPSGAS 1240
```

```
QY      583 PSETSKLOQLVEKIDROGAVAVASTASGAPTTAP---APSSS-ASGPNOCV---ICLRVL 636
      1241 PSASALTILGL-----ATAPSLSSSQTPPGHPLLPATTSRHVGLNSTVAPACSPVL 1290
OY      637 SCPRALRHYGQHGERPFFKCKVCGRAFSTRGNLRNHFHGHKTSPPARAQNSCPICQKF 696
      1291 VPASAL-----ASFPFAPNPAPA----- 1309
OY      697 TNAVTLQOHVRLMLGGQIPNGGSALSEGGAQENSSEOSTASGSPSPFOQOSQPSPEE 756
      1310 -----QASLLAPASASASQ-----ALATPLPMAAPQITALIAPSPAP 1345
OY      757 EMSEEEEDDEEBEDVDDEDSLARGSESQCKAISVRDSEVYSAEEVATSVAAAA 812
      1346 PLA-----PLPVLPAPSPGAAPVAVLASQTPVPVMAPSSTPGTSLASASP 1388
OY      813 --APTTVKEMDSNEKARQHTLPPPPPPPDNLDPQPMEOGTSVDSGAMBEAKLEGISSP 870
      1389 VPAPTPLYLAPSSST---QTMLPAPVPSF---LPSPASTQTALALAPAL---APTGGSSP 1437
OY      871 MAALTOEBEGSTPPL-----VEELNLPEA-----MKKDPGESSGKRACEVCGSFP 916
      1438 SQTLSL---GTGNPQGGFPPTQTLSTLPASSLVPTRPAQTLSLAPGPLG-----P 1483
OY      917 TOTALBEOHKTTHPKDGLFTCVFCQGFIDRATLKKHMLLAHHQVPRFAPH---GPONIA 973
      1484 TOT-----LSLAPADPLAPASVPVGAAPAH 1507
OY      974 TSLVPGCSS-----SIPSPGLSPFP 994
      1508 TLTLPAPASSASASLAPASVQTLTLPAP 1535

RESULT 26
US-09-579-181-1
: Sequence 1, Application US/09579181
: Patent No. 6365372
: GENERAL INFORMATION:
: APPLICANT: Chirivia, John
: APPLICANT: Yaciuk, Peter
: TITLE OF INVENTION: SNE2 Related CBP Activator Protein (SRCAP)
: FILE REFERENCE: 16153-4247
: CURRENT APPLICATION NUMBER: US/09/579,181
: CURRENT FILING DATE: 2000-05-25
: PRIOR APPLICATION NUMBER: 60/136,620
: PRIOR FILING DATE: 1999-05-27
: NUMBER OF SEQ ID NOS: 17
: SOFTWARE: Patentl Ver. 2.0
: SEQ ID NO 1
: LENGTH: 3118
: TYPE: PRT
: ORGANISM: Human
: US-09-579-181-1

Query Match      4.2%; Score 224.5; DB 4; Length 3118;
Best Local Similarity 20.8%; Pred. No. 2.1e-06;
Matches 218; Conservative 78; Mismatches 317; Indels 435; Gaps 42;

OY      72 NSSASSAPRPEGHSRQVMDTEHSNPPDSGSGPPDPPTWGPERRGESSGQFLVAATGTA 131
      944 NMLQPVPRKQEGRTVVVV-----NNPRAPLGPVVRPPPGPELSAQPTGPPV----- 990
OY      132 AGGGGGLILASPRLGATPLPPESTPAPPPPP-----PPPPPGVSGHLNPLIL-EBLRVL 187
      991 ----POVLASLSLWVSASPAGPRLIPASRPBPVLLPILQPNNGS----LPQVLSPPLGVL 1042
OY      188 QORQIHQMOMTEQICQOVLLLSLQGTVGAPASPSSELPGTGAASKRLLPLFSPPIKPAQ 247
      1043 SGTSS-----RPPPTLSLKPFPAPVRLSPAPPPGSSSLKPL-----TYPPGY 1086
OY      248 T-----GKTTASSSSSSSSSGAEPKQAFPHLYHPLGSHQHPFSVGVGGRSHKPTPASPAL 303
```


1087 TPPAATTTTATTATTAVP-----APTAP----- 1115
304 PGSTDOLASPHLAPPTGTGLLAOCLGARCLGAASPGILKRNKSGE-LGYEVYIS 362
1116 ---ORLILSPD-----MOARLPSGEVYISIQOLASL 1142
363 LEKPGGHHKRCFCAKVFSGDSALQILRSHGTGERPYKCNVGNFTTGG-----NLKVEHH 418
1143 AORP-----VANAGGSKPLTFQIOGNKLTLLTGAOYROLAVGCP 1180
419 RHREKYPHVQMN-----PHVPEHLDYVITSSGLPYGMSVPP 455
1181 RPLQMPETPMVNNNTGVVKTIVVROAPRDGLTPVPELAPAPRP-----PSSGHPAVILNPPRP 1233
456 EKAEDEEGTGGGVEKRLVASTALSAATESLTLSTGTSTAVAPGLPTFNKFLMKAVE 515
1234 TLPPGLRPLPLGTARABM-----PTPLVLRP-----LTKLVH 1266
516 PKSKADENTPPGSE-----GSAIAGVADS-----GSATRMOLS----- 548
1267 SPSPREVSASAPGARPLTITSSPLHVPSSSLPFGPASSPMPITPNSPLASPVSTYVSPILSSSL 1326
549 --KLVTSLPS-----WALLTNHLKSTG-----SPFPYVLEPLGAS 582
1327 PISVPTPLPAPASAPLTIPIAPLTVSASGAPLITSVTPPLAPVAPAPGPPSLQPSGAS 1386
583 PSETSKIOQLVEKIDROCAVAVASTASGAPTTAP---APSSS-ASGPNOCY--ICLRVL 636
1387 PSASALATLGL-----ATAFSLSSQTPGHPHLLAPLSSHVPGINSTVAPACSPVL 1436
637 SCPRALRHYGHGGERPFKCKVCGRAPFTRGNLRAHFVGHKTSPPARAONSCPTQKKF 696
1437 VPASAL-----ASPPSAPNAPR----- 1455
697 TNAVTLQOHVHMLGGQIPNGSALSSEGGGAOENSSSEOSTASGSPGSPPOSGOPSPPEE 756
1456 -----QASLLAPASSAQ-----ALATPLAPMAQTAIILASPAP 1491
757 EMEDEEEDDEEDVTDSDLAGRSGESGGEKAIIVRGDSEEVSGAEVATSA----- 812
1492 PTLA-----PPLVLAAPSPGARVYLASSQTPVPVWAPASSTPCTSLASASP 1534
813 --APTVKEMDSNEKAPONTLPPPPPPNDLHPQPMEOGSTDVSGAMEEAKLEGISSP 870
1535 VPAPPTVPLAPBST---QTMPLAPVPSR---LPSPASTQTLALAPAL---APTLGSSSP 1583
871 MAALTOGEGISTPL---VEELNLPFA-----MKDPESSGRKACEVCGQSP 916
1584 SQTLSL---GTGNPQGPPTQTLISLPASSLVPTPAOTLSLAPGPPILG-----P 1629
917 TOTALERHOKTHPKDGPLFTVCFCRQGFIDRATLKKHMLLAHQVPPAPRH---GPQNTA 973
1630 TQI-----LSLAAPPLAPASPVGAPAPAH 1653
974 TLSLVPGCSS-----SIPSPGLSPFP 994
1654 TLFLLAPASSASLAPASVQTLTSLSPAP 1681

RESULT 27
JS-08-102-942A-4
Sequence 4, Application US/08102942A
Patent No. 5726288

GENERAL INFORMATION:

APPLICANT: Call, Katherine M.
APPLICANT: Glaser, Thomas M.
APPLICANT: Ito, Caryn Y.
APPLICANT: Buckler, Alan J.
APPLICANT: Pelletier, Jerry
APPLICANT: Haber, Daniel A.
APPLICANT: Rose, Elise A.
APPLICANT: Housman, David E.
APPLICANT: Bruening, Wendy

APPLICANT: Darveau, Andre
TITLE OF INVENTION: Localization and Characterization of the
TITLE OF INVENTION: Wilms' Tumor Gene
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESS: Hamilton, Brook, Smith & Reynolds, P.C.
STREET: Two Militia Drive
CITY: Lexington
STATE: Massachusetts
COUNTRY: U.S.
ZIP: 02173
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/102, 942A
FILING DATE: 02-AUG-1993
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Granahan, Patricia
REGISTRATION NUMBER: 32,227
REFERENCE/DOCKET NUMBER: MIT-5194A2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-861-6240
TELEFAX: 617-861-9540
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 449 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-102-942A-4

Query Match 4.2%; Score 222; DB 1; Length 449;
Best Local Similarity 25.6%; Pred. No. 2.5e-07;
Matches 100; Conservative 38; Mismatches 135; Indels 118; Gaps 20;

QY 124 LVNATGTAAGGGGLIILASPKLGA---TPL-----PPEST-----PAPPPPPPPPPPG 169
DB 11 LPLAVSSLLGGGGGCGGL-PVSGARQMAPVLDPAFGASAYGSLGGPAPPPAPPPPPPP 67
QY 170 VCSGHNLIPLLEELRYLQCRQIHOHOMTEQICRGVLLGSLGQ---TVGA-----PA 219
DB 68 -----PHSFIKQEPSWGAEPHEEQCLSAFTLHFSQFTGTAGACRQGFPP 115
QY 220 SPSEL-PTGAASSTKPLPLPLSPKPA--GTGKTTA-----SSSSSSSSGAEPPKQA 270
DB 116 PPSQASSGQARMFPNAPVLPSCLESQPTIRNOGYSTVTEPDGAPSYGHTPSHHAQFPNHS 175
QY 271 FPHLYHPILGSOHPFSVGVGRSHKPTP-----APSPALPGSDOLIASPHLA---PPG 320
DB 176 FKH-EDPMGQD---SLSGEOQYSVPPPVYCGHTTDSCTGSOALLIRTPYSSDNLTYOM 229
QY 321 TTGLLAQC-----LGAARGLAASAPGLPKNGSGELGYGEVYISLEKP---GGRH 370
DB 230 TSQL---ECMTWNOGNLIGATLKGMAAGSSSVKWTGQSNHGGEYSENHAPLILCGAQY 286
QY 371 K-----CRF---CAKVFSGDSALQIHLRS 391
DB 287 RIHTHGVFRGIQDVRVRSVAPTLVRSASETSEKRPDMCAVPGCKNRYFKLSHLOMHSRK 346
QY 392 HNGERPYKCNV--CGNRTTGNLKVHFRH 420
DB 347 HNGEPYQCDKCDERRRSRSDQLKRHORH 377

RESULT 28
US-09-037-179B-4
Sequence 4, Application US/09037179B
Patent No. 6316599
GENERAL INFORMATION:

APPLICANT: Call, Katherine M.
APPLICANT: Glaser, Thomas M.
APPLICANT: Ito, Caryn Y.
APPLICANT: Buckler, Alan J.
APPLICANT: Pelletier, Jerry
APPLICANT: Haber, Daniel A.
APPLICANT: Rose, Elise A.
APPLICANT: Hausman, David E.
APPLICANT: Bruening, Wendy
APPLICANT: Darveau, Andre
TITLE OF INVENTION: Localization and Characterization of the
FILE OF INVENTION: Wilms' Tumor Gene
FILE REFERENCE: 0050.1312-011
CURRENT APPLICATION NUMBER: US/09/037,179B
CURRENT FILING DATE: 1998-03-09
PRIOR APPLICATION NUMBER: US 08/102,942
PRIOR FILING DATE: 1993-08-02
PRIOR APPLICATION NUMBER: US 07/614,161
PRIOR FILING DATE: 1990-11-13
PRIOR APPLICATION NUMBER: US 07/435,780
PRIOR FILING DATE: 1989-11-13
PRIOR APPLICATION NUMBER: US 07/795,323
PRIOR FILING DATE: 1994-09-27
NUMBER OF SEQ ID NOS: 21
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ ID NO 4
LENGTH: 449
TYPE: PRT
ORGANISM: Unknown
FEATURE:
OTHER INFORMATION: Murine
US-09-037-179B-4

Query Match 4.2%; Score 222; DB 4; Length 449;
Best Local Similarity 25.6%; Pred. No. 2.5e-07;
Matches 100; Conservative 38; Mismatches 135; Indels 118; Gaps 20;

124 LVAATGTAAGGGGGLILASPKLGA--TPL-----PPEST-----PAPPPPPPPPPG 169
11 LPPAVSSLGSGGGGCGL--PVGARQMAPIVLDFAIPGASAGSLGPPAPPPPPPPPP 67
170 VGSGLNPLILEELRVLDQRIHQOMTEQICROYLLGSLGQ--TVGA-----PA 219
68 -----PHSFIKQEPKPSWGAEPHEQCISAFILHFSQFTGAGACRYGPPGPP 115
220 SPSEL-PGTGAASSTKPLPLPFSPIKA--QTGKTA-----SSSSSSSSGAEPPKQA 270
116 PPSQASSQQAHPFAPRYPLPSCLESQPTIRNOGISTVTFDGAIPSYGHTPSHHAQFPNHS 175
271 FEHLVHPLGSHPFVSGVGRSHKPTP-----APSPALPGSTDLQILASPHLA--FPG 320
176 FKH-EDPWGQQ-----GSLGEQOYSVRPVYGGCHTPDSCGSQLALLRTYSSDNLYOM 229
321 TTGLLAQC-----LGAARGLAASPELLPRKNGSGELGYEVSSLEKP---GGRH 370
230 TSQL-----ECMTWNOXNLGATLTKGMAAGSSSVKWTGQSNHGTGYESENHTAPILCGAQY 286
371 K-----CRP-----CAKYGCSALQIHLRS 391
287 RIHTHGVFRGIQDVARYSVAPTLVNASASETSEKRRPMCAIPGCKNKRTFKLSHLMHNRK 346
392 HTGERPKYKCNV--CGNRFTTNGNLKLVHFRH 420
347 HTGEKPYOCDFKDCERFRSRSQDLKRKHRRH 377

RESULT 29
US-08-224-482-2
Sequence 2, Application US/08224482
Patent No. 5837692
GENERAL INFORMATION:
APPLICANT: Mercola, Dan
APPLICANT: Adamson, Eileen D.

TITLE OF INVENTION: Inhibition of the Mitogenic Activity of
TITLE OF INVENTION: PDGF by Mammalian EGR
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Campbell and Flores
STREET: 4370 La Jolla Village Drive, Suite 700
CITY: San Diego
STATE: California
COUNTRY: USA
ZIP: 92122
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/224,482
FILING DATE: 07-Apr-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-ME 9913
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 496 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-224-482-2

Query Match 4.2%; Score 220.5; DB 2; Length 496;
Best Local Similarity 22.5%; Pred. No. 3.6e-07;
Matches 120; Conservative 59; Mismatches 179; Indels 175; Gaps 25;

97 PPDSSGSPDDPTWGPBRNGEESGQFLVAATGTAAGGGGGLILASPKLGAATPLPPESTP 156
9 PQLFGAAGTP-----EGSGGNSSTSSGSGGGGG-----SNAGSSAFNPGEP 52
157 APPPPPPPPPPGVSGHLNPLILEELRVLDQRIHQOMTEQICROYLLGSLGQYVG 216
53 SHQP-----YELITE--SFSDIALNNEKAMETSYPSQYTR---LPPIYTR 96
217 APASPELPGTGAASSTKPLPLPFSPI-----KPAOTGKTASSSSSSSSSGAEPPQAFF 272
97 FSLERPA--PNSGNTLMPPLFSLVSGVSMINPPTSSSASFPASSSSSSASQSP----- 149
273 HLYHPLGSHPFVSGVGRSHKPTP-----SPALPGSTDLQILASPHLAEPGTT 322
150 ---PLSCAVPNSDSSPIYSAAETFPPTPNDIFPEPQSAFPESAGTALQYPPPAVPATK 205
323 G-----LLAAQ-----CLGAA-----RGLEAASPGLLKP-----KNGSGEL 354
206 GGFQVMPIDYLPFOOQGLDLSLSTPDOKPFGLENRTOQPSLTPLSTIKAFATQSGSODL 265
355 GYGEVI--SLEKPG-----GRH----- 370
266 KALNTTYQSOLIKPSMRKYVPNRPKTPRHERPYACVESCDRRSDELTNRHRIHTG 325
371 ---KCRCAKAVFGSSALQIHLRSHTGERPYKCNVCGNRFT---TRGNLKVHFRHRE 422
326 QKPFQCRICMRNFSRSDHLTTHRTHTGEKPFACDICGRKPARSDERKRNHTKILHROKXK 385
423 K-----YPHVQMPHPVREHLDY---VITSSGLPYGMSVPEKAELENG 463
386 KADKSVASPAASSLSSTYSPTATSTPSPATTSPPSPVPTSYSSP--GSSTYPSPA--HSG 442
464 TPGGGERKPLVAST-----TALSATESLTLST--GTSAVAPDLPTFN 506
443 FP-----SPSVATTFASVPAPFPTQVSSPSPSAGVSSSFSTGLSDMTATFS 489

RESULT 30
S-08-234-783-4
Sequence 4, Application US/08234783
Patent No. 5622835
GENERAL INFORMATION:
APPLICANT: Herlyn, Meenhard
APPLICANT: Morris, Jennifer
APPLICANT: Rauscher III, Frank J.
APPLICANT: Rodeck, Ulrich
TITLE OF INVENTION: Wt1 Monoclonal Antibodies and Methods of
TITLE OF INVENTION: Use Therefor
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: Howson and Howson
STREET: Spring House Corporate Cntr, PO Box 457
CITY: Spring House
STATE: Pennsylvania
COUNTRY: USA
ZIP: 19477
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/234,783
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Bak, Mary E.
REGISTRATION NUMBER: 31,215
REFERENCE/DOCKET NUMBER: WST48USA
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-540-9200
TELEFAX: 215-540-5818
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 429 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
S-08-234-783-4
Query Match 4.1%; Score 217; DB 1; Length 429;
Best Local Similarity 25.1%; Pred. No. 5.1e-07;
Matches 94; Conservative 38; Mismatches 123; Indels 120; Gaps 18;
133 GGGGGLIASFKLGATP-----LPREST-----PAPPPPPPPPPGVSGHLNIP 178
19 GGGGGCAL--PVSQAQAQAPVLDFAFPAGASAYGSLGAPAPPPAPPPPPPHS----- 70
179 LILEELRVLOQRQHOMQTEQICROVLLGSLGQ---TWGA-----PASPSEL-PGT 227
71 -----FIKQPSWGAERHEOCLSAFYVHFGQFTGTAGACRYGPRPPPSQASSG 124
228 GAASSTPRLPLFSPRIKA-----QTGKTTAASSSSSSSGAEPKQAFPHLYHPLG 279
125 ARMEPNAPYLPSCLESQPAIRFGYSTVFEDGTPSYGHTPSHHAQFPNHSFKH-EDPMG 183
280 SQHPFSYGVGVRSHKPTP-----APSPALPGSTQOLIASPHLA---FPGTTGL----- 324
184 QO-----GSLGEQOYSVPPVYGCHPTDSCGTGQALLRTPYSSDNLXQMTSOLECMTW 238
325 -----LAAQCLGAARGLA-----AASPGILPKPN 349
239 NQANLGAFLKGHSTYESNHTPIILGAYQRIHTHGVRFGIQRVRYGVAPFTLYRSAS 298
350 GSGELVGEVLSLEKPGGRHRCRF--CAVFGSDASALQIHLRSHTGERPYCNV--CGN 405
299 ETSE-----KRP--FMCAVPGCNKRYFKLSHLQMSRKHTGKERPYOCDPDCER 345

QY 406 RETTRGNLKVHFHRH 420
DB 346 RFRSDQLKRHORH 360
RESULT 31
US-08-456-907-4
Sequence 4, Application US/08456907
Patent No. 5633142
GENERAL INFORMATION:
APPLICANT: Herlyn, Meenhard
APPLICANT: Morris, Jennifer
APPLICANT: Rauscher III, Frank J.
APPLICANT: Rodeck, Ulrich
TITLE OF INVENTION: Wt1 Monoclonal Antibodies and Methods of
TITLE OF INVENTION: Use Therefor
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: Howson and Howson
STREET: Spring House Corporate Cntr, PO Box 457
CITY: Spring House
STATE: Pennsylvania
COUNTRY: USA
ZIP: 19477
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/456,907
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/234,783
FILING DATE: 28-APR-1994
ATTORNEY/AGENT INFORMATION:
NAME: Bak, Mary E.
REGISTRATION NUMBER: 31,215
REFERENCE/DOCKET NUMBER: WST48AUSA
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-540-9200
TELEFAX: 215-540-5818
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 429 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-456-907-4
Query Match 4.1%; Score 217; DB 1; Length 429;
Best Local Similarity 25.1%; Pred. No. 5.1e-07;
Matches 94; Conservative 38; Mismatches 123; Indels 120; Gaps 18;
133 GGGGGLIASFKLGATP-----LPREST-----PAPPPPPPPPPGVSGHLNIP 178
19 GGGGGCAL--PVSQAQAQAPVLDFAFPAGASAYGSLGAPAPPPAPPPPPPHS----- 70
179 LILEELRVLOQRQHOMQTEQICROVLLGSLGQ---TWGA-----PASPSEL-PGT 227
71 -----FIKQPSWGAERHEOCLSAFYVHFGQFTGTAGACRYGPRPPPSQASSG 124
228 GAASSTPRLPLFSPRIKA-----QTGKTTAASSSSSSSGAEPKQAFPHLYHPLG 279
125 ARMEPNAPYLPSCLESQPAIRFGYSTVFEDGTPSYGHTPSHHAQFPNHSFKH-EDPMG 183
280 SQHPFSYGVGVRSHKPTP-----APSPALPGSTQOLIASPHLA---FPGTTGL----- 324
184 QO-----GSLGEQOYSVPPVYGCHPTDSCGTGQALLRTPYSSDNLXQMTSOLECMTW 238
325 -----LAAQCLGAARGLA-----AASPGILPKPN 349

Db 239 NOMNLGATLKSHSTGYESDNHTTPIILCAQYRIHTHGVFRGIQDVRVPGVAPTLVRSAS 298
OY 350 GSGELGYGEVSISSLEKPGRHKCRF--CAKVFSGDSALQIHLRSHTEGPRYKCNV--CGN 405
Db 299 ETSE-----KRP---FMCAVPGCNKRYFKLSLQMSHRKHTGEPYOCDFKDCER 345
OY 406 RFTTRGNLKVHFHRH 420
Db 346 RFSRSDQLKRHRH 360

RESULT 32

PCT-US95-05523-4
Sequence 4, Application EC/TUS9505523
GENERAL INFORMATION:
APPLICANT: The Wistar Institute of Anatomy and Biology
TITLE OF INVENTION: M1 Monoclonal Antibodies and
METHODS OF USE THEREFOR
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Howson and Howson
STREET: Spring House Corporate Cntr. PO Box 457
CITY: Spring House
STATE: Pennsylvania
COUNTRY: USA
ZIP: 19477
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/05523
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/234,783
FILING DATE: 28-APR-1994
ATTORNEY/AGENT INFORMATION:
NAME: Bak, Mary E.
REGISTRATION NUMBER: 31,215
REFERENCE/DOCKET NUMBER: WST488PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-540-9200
TELEFAX: 215-540-5818
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 429 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US95-05523-4

Query Match 4.1%; Score 217; DB 5; Length 429;
Best Local Similarity 25.1%; Pred. No. 5.1e-07;

Matches 94; Conservative 38; Mismatches 123; Indels 120; Gaps 18;

OY 133 GGGGGLILASFKLGATP-----LPPEST-----PAPPPPPPPPPPGVSGSHLNP 178
Db 19 GGGGGCAL--PVSGAAMNAPVDFAPRGASAVYSLGSPAPPPAPPPPPPHS----- 70
OY 179 LILELRLVLRQRIHQMTEDICROYLLGLSGQ--TWGA-----PASPSEL-PGT 227
Db 71 -----FLIKQPSWCGAHEHCISAFIVHFSGQFTGTAGACRYGPPGPPPSQASSSQ 124
OY 228 GAASSTKPLPLFESFKR-----OTGKTASSSSSSSSGAEPKQAFHLIYHPLG 279
Db 125 ARMRPNAPYLPSCLSEQRAIRINQGVSTVFDGTPSYGHTPSHHAQEPNHSFKR-EDPVG 183
OY 280 SQAPFSYGVGVSRSHKPTP-----APSPALPGSTDOILASPHLA--FPGTTGL----- 324
Db 184 QQ-----GSLGEQGVSVPPVYGCHTPTDSCSTGSQALLLRTFPYSSDNLYQMTSQLECMTW 238

OY 325 ----LAAQCLGARGLEA-----AASPGLLKPN 349
Db 239 NOMNLGATLKSHSTGYESDNHTTPIILCAQYRIHTHGVFRGIQDVRVPGVAPTLVRSAS 298
OY 350 GSGELGYGEVSISSLEKPGRHKCRF--CAKVFSGDSALQIHLRSHTEGPRYKCNV--CGN 405
Db 299 ETSE-----KRP---FMCAVPGCNKRYFKLSLQMSHRKHTGEPYOCDFKDCER 345
OY 406 RFTTRGNLKVHFHRH 420
Db 346 RFSRSDQLKRHRH 360

RESULT 33

US-08-246-489-2
Sequence 2, Application US/08246489
Patent No. 6225049
GENERAL INFORMATION:
APPLICANT: Ian, Michael S.
TITLE OF INVENTION: NOVEL HUMAN INSULINOMA-ASSOCIATED CDNA
NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Knobb, Martens, Olson & Bear
STREET: 620 Newport Center Drive
CITY: Newport Beach
STATE: California
COUNTRY: USA
ZIP: 92660
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/246,489
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/901,715
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Israelson, Ned A.
REGISTRATION NUMBER: 29,655
REFERENCE/DOCKET NUMBER: NIH012.012A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 510 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: N-terminal
US-08-246-489-2

Query Match 4.1%; Score 216.5; DB 4; Length 510;
Best Local Similarity 22.7%; Pred. No. 6.9e-07;
Matches 140; Conservative 46; Mismatches 229; Indels 201; Gaps 30;

OY 133 GG--GGGGLILASFKLG--ATPLPESTPAPPPPPPP-----PPPP 168
Db 22 GGEDDDRALLLSPSCGABAREPAPSPVPPPLPPPPAERAAHAAALAACAPGPQPPO 81
OY 169 GVGSGHLN-----IPLILELRLVLRQRIHQMTEDICROYLLGLSL 211
Db 82 GPRAAHFGPPEAHPARPLSPTRPVGREHEKHKYFERSFNLSYSAESFPRAALLGG 141
OY 212 GQYGAAPASPSSELPGTAASSTKPLPLFSPTRPAQGTGTTASSSSSSSGCAE-----P 266

142 G---GGGASGA---GGGTCGGDPLL-----FAPRLKMGTAFSAGAFAARGPGRPLP 190
Y PROAFPHLYHPLSGQHFFSVSGVGRSHKPTAPSPALPGSTDOIAPSHLAPGTTGLIA 326
Y 191 PAAAL---RPPGRKRP-----PPTAAEPRA-----KAVKAPGAKKP----- 223
Y 327 AAGLGAARGL---EAAASPL-LKPKNGSGELGYGEVISLSLEKPRGRHKCRFCAKVGS 381
Y 224 ---KAIKRLHFEDEVTTSPVLGIKIKGVEAPRCRA-GGAAPRLGELTICQLKEEYAD 278
Y 382 DSALQIHLRSHTEGPRYKCNVCGNFRFTTGRNLKVFHREKRYHVOMNPHRPHLDYV 441
Y 279 PRALAQHKCRIVAVEYRCPECAKVFSCPANLASHRMK-----PRPAP----- 323
Y 442 ITSSGLPYGMSVPEKA---EEBAGTPGGGVKRRPLVASTALATESITLITSTSTAVA 499
Y 324 -AAARAP-----EPEAARAEAREAPGG----- 346
Y 500 PGLPTFKFVLMKAVEPKSKADENPR-PGSEGSALAGVADSGSATRM-QLSKLYTSLPSW 557
Y 347 -----SDRDTPPSPG-----GVSESGSEDGLYECHHCAKFRRO 379
Y 558 ALLTNH-----LKSTGSPFPYVLEPLGASPSMTSKLOQLYVEKIDROGAVAVASTASG 610
Y 380 AYLRKHLAHQALQAKA-----PL-APPAP--DLALYGPDEKADQEAAGDGG 428
Y 611 APTTAPAPSSASGPNOCVICTLRYLSCPRALRLHYGQHGGERPFCKVCYCGRAESTRGL 670
Y 429 AGVLGL-----SASAECHLCPYCGESFASKGAEERHLRLHAAQVFPCKYCATFYSSPL 484
Y 671 RAHFVGHKTSAPARAQ 686
Y 485 TRHI--NKCHPSENRQ 498

RESULT 34
US-08-102-942A-6
Sequence 6, Application US/08102942A
Patent No. 5726288

GENERAL INFORMATION:

APPLICANT: Call, Katherine M.
APPLICANT: Glaser, Thomas M.
APPLICANT: Ito, Caryn Y.
APPLICANT: Buckler, Alan J.
APPLICANT: Pelletier, Jerry
APPLICANT: Haber, Daniel A.
APPLICANT: Rose, Elise A.
APPLICANT: Housman, David E.
APPLICANT: Bruening, Wendy
APPLICANT: Darveau, Andre
TITLE OF INVENTION: Localization and Characterization of the
TITLE OF INVENTION: Wilms' Tumor Gene
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
STREET: Two Militia Drive
CITY: Lexington
STATE: Massachusetts
COUNTRY: U.S.
ZIP: 02173

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/102,942A
FILING DATE: 02-AUG-1993

CLASSIFICATION: 530

ATTORNEY/AGENT INFORMATION:
NAME: Granahan, Patricia
REGISTRATION NUMBER: 32,227
REFERENCE/DOCKET NUMBER: MIT-5194A2

TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-861-6240
TELEFAX: 617-861-9540
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 449 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-102-942A-6

Query Match 4.1%; Score 215.5; DB 1; Length 449;
Best Local Similarity 25.4%; Pred. No. 6.9e-07;
Matches 97; Conservative 35; Mismatches 133; Indels 117; Gaps 16;

QY 133 GGGGGLILASPKLGATP-----LPPEST-----PAPPPPPPPPPGVSGHLNIP 178
DB 19 GGGGGGAL--PVSQAQWAPYLIDFAPBPASAYGSLGPAAPPAPPPPPPHS----- 70
QY 179 LILIELRVLDQRQIHQOMTQICRQVILLGSLGQ--TVGA-----PASPEL-PGT 227
DB 71 -----FTRQEPMSWGAEPHEQCISAFTHPSGQFTGTAGACRYGPFPPPSQASSGQ 124
QY 228 GAASSTKPLPLFSPKIPA-----OTGKTTASSSSSSSSGAEPKQAFPHLYHPLG 279
DB 125 ARNFPNAPRYLPSCLESQPALNNGYSTVTFDGTPSYGHTRPSHHAQRPNNSEKH-EDPMG 183
QY 280 SQHPFSVGVGSRSHKPRP-----APSPALPGSTDOIAPSHLA---FPGTTGILAAQC 329
DB 184 QQ-----GSLGEQGYVPPRYGCHTPDSCTSQALMLTTPYSSDNLQYMTSQL--EC 235
QY 330 -----LGAARGLEAASPLILKPKNGSGELGYGEVISLEKP---GGRHK----- 371
DB 236 MTWNQNLGATLKGVAAGSSSVKWTBEGOSHNSTGYESDNNHTPIILCAQYRITHGYFR 295
QY 372 -----CRF--CAKVGSDSALQIHLRSHTEGPRYK 400
DB 296 GIQDVRVPGVAPTLVVSASSETSEKRPFCACPGCKRRFKLHLQHSRKHTEGKRYQC 355
QY 401 NV--CGNRFETTRGNLKVHFRH 420
DB 356 DFKDCERRFSRSDQLKRORRH 377

RESULT 35
US-09-037-179B-6
Sequence 6, Application US/09037179B
Patent No. 6316599

GENERAL INFORMATION:
APPLICANT: Call, Katherine M.
APPLICANT: Glaser, Thomas M.
APPLICANT: Ito, Caryn Y.
APPLICANT: Buckler, Alan J.
APPLICANT: Pelletier, Jerry
APPLICANT: Haber, Daniel A.
APPLICANT: Rose, Elise A.
APPLICANT: Housman, David E.
APPLICANT: Bruening, Wendy
APPLICANT: Darveau, Andre
TITLE OF INVENTION: Localization and Characterization of the
TITLE OF INVENTION: Wilms' Tumor Gene
FILE REFERENCE: 0050.1312-011
CURRENT APPLICATION NUMBER: US/09/037,179B
CURRENT FILING DATE: 1998-03-09
PRIOR APPLICATION NUMBER: US 08/102,942
PRIOR FILING DATE: 1993-08-02
PRIOR APPLICATION NUMBER: US 07/614,161
PRIOR FILING DATE: 1990-11-13
PRIOR APPLICATION NUMBER: US 07/435,780
PRIOR FILING DATE: 1989-11-13
PRIOR APPLICATION NUMBER: US 07/795,323
PRIOR FILING DATE: 1994-09-27


```

: NUMBER OF SEQ. ID NOS: 21
: SOFTWARE: FASTSEQ for Windows Version 4.0
: SEQ ID NO 6
: LENGTH: 449
: TYPE: PRT
: ORGANISM: Homo sapien
US-09-037-179B-6

Query Match      4 18; Score 215.5; DB 4; Length 449;
Best Local Similarity 25.4%; Pred. No. 6.9e-07;
Matches 97; Conservative 35; Mismatches 133; Indels 117; Gaps 18;

QY 133 GGGGGLILASPKIGATP-----LPPEST-----PAPPPPPPPPPGVSGHLNIP 178
      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 19 GGGGGGAL--PYSGAQMNAVLDPAFPGASAYGSLGGRAPAPAPPPPPPPPHS----- 70
      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 179 LLEELRVLQQRQIOMOMTEQICROYLLGSLGQ---TVGA-----PASPEL-PGT 227
      : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 71 -----FTKQEPSWGGAEPHEQOLSAFTVHFSGGFTGAGACRYGPFPGPPPSQASSGQ 124
      : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 228 GAASSTKPLPLPESPICKPA-----QTGKTTASSSSSSSSSGAEPKQAFILYHPLG 279
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 125 ARMPNAYLPSPCLEQPAIRNCGYSTVTFPGTPTGYGHTPSHIAQFPNHSFKH-EDMG 183
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 280 SQHPFVSGVGRSHKPTP-----APSPALPGSTDLIASPHLA--PFGTGLLAQC 329
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 184 QQ-----SLSGQQYQVSPRPVYGCHTPTDCTGSOALLLRTPTPSDNLXOMTSOL--EC 235
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 330 -----GGAAGGLEAASPGLLKPKNSGELGYEVISLEKPP--GGRHK----- 371
      ||||| | | | | | | | | | | | | | | | | | | | | | | |
DB 236 MTWNOMNLGATLKGVAAAGSSSVAKWTGOSNHSTGYESDNNHTPTILCGAOTRIRITHGVFR 295
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 372 -----CRP--CAKVGSDSALQIHLRSHTGERPYKC 400
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 296 GIDVPRVPGVAPTLVRSASETSEKRPFCMCAYPGCMKRYFKLSHLQMSRKHGTGKPYOC 355
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 401 NV--CGNRFTTGNLKVHFHRH 420
      : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 356 DFKDCRRRSRSDQLKRHQRHH 377
      : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

RESULT 36
US-08-545-860D-55
: Sequence 55, Application US/08545860D
: Patent No. 6040140
: GENERAL INFORMATION:
: APPLICANT: Croce, Carlo
: APPLICANT: Canaan, Eli
: TITLE OF INVENTION: Diagnostics, Therapeutics and Methods
: FOR DETECTION AND TREATMENT OF ACUTE LEUKEMIAS
: TITLE OF INVENTION: Resulting from Chromosome Abnormalities in the ALL-1 Region
: NUMBER OF SEQUENCES: 94
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Woodcock, Washburn, Kurtz, Macklewitz &
: ADDRESS: No. 6040140r1s
: STREET: One Liberty Place, 46th floor
: CITY: Philadelphia
: STATE: Pennsylvania
: COUNTRY: USA
: ZIP: 19103
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/545, 860D
: FILING DATE: 07-MAR-1996
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: PCT/US94/04496
: FILING DATE: 22-APR-1994
: PRIOR APPLICATION DATA:

```

	APPLICATION NUMBER:	PCT/US92/10930	
	FILING DATE:	09-DEC-1992	
	PRIOR APPLICATION DATA:		
	APPLICATION NUMBER:	US 08/327,392	
	FILING DATE:	19-OCT-1994	
	PRIOR APPLICATION DATA:		
	APPLICATION NUMBER:	US 08/320,559	
	FILING DATE:	11-OCT-1994	
	PRIOR APPLICATION DATA:		
	APPLICATION NUMBER:	US 08/062,443	
	FILING DATE:	14-MAY-1993	
	PRIOR APPLICATION DATA:		
	APPLICATION NUMBER:	US 07/971,094	
	FILING DATE:	30-OCT-1992	
	PRIOR APPLICATION DATA:		
	APPLICATION NUMBER:	US 07/888,839	
	FILING DATE:	27-MAY-1992	
	PRIOR APPLICATION DATA:		
	APPLICATION NUMBER:	US 07/805,093	
	FILING DATE:	11-DEC-1991	
	ATTORNEY/AGENT INFORMATION:		
	NAME:	Deluca Esq., Mark	
	REGISTRATION NUMBER:	33,229	
	REFERENCE/DOCKET NUMBER:	TJU-1262	
	TELEPHONE:	(215) 568-3100	
	TELEFAX:	(215) 568-3439	
	INFORMATION FOR SEQ ID NO:	55:	
	SEQUENCE CHARACTERISTICS:		
	LENGTH:	1093 amino acids	
	TYPE:	amino acid	
	STRANDEDNESS:	single	
	TOPOLOGY:	linear	
	MOLECULE TYPE:	protein	
	HYPOTHETICAL:	NO	
	ANTI-SENSE:	NO	
	US-08-545-860D-55		
Query Match	4.0%;	Score 211.5;	DB 3; Length 1093;
Best Local Similarity	20.6%;	Pred. No. 4.1e-06;	
Matches 192;	Conservative 88;	Mismatches 321;	Indels 329; Gaps 41.
QY	3 QEFGSSSRILGPGCEPAPRGDASEEHHPDYOVAKCAOFSDPEFLAHONSCCTDPYMW	62	
DB	289 QETSSESR-----ESKSKKKSSHLSLHKCKLLSGCKGVSTTSASSSSSSSSS-	338	
QY	63 IIGQENPNSSNASSAPPREGHSH-SQVMDTEHNPDDSGSGPPDT-----	109	
DB	339 --GGPFQPAVSSLSPDSFSAFKLEQPEDEDKSKPTAPAPASAPSPSAPPRADELFEQ	396	
QY	110 -----WCPER--GESSSGOFLVAATG-----TAAGGGGLLASPKLGATPLPE	153	
DB	397 KVFPSGGPIRPFSTTTTSSSGRARAPSPEDYKSPHYTGSGASACTHKMPALSKTPPAD	456	
QY	154 STP-----APP--APP--PPPPPG-----	169	
DB	457 ETPEPTGLKEKKHKASKRSRHGRPRKSGRNKEGTGCPAAPSLSQAOLAGFTATAASPESG	516	
QY	170 ---VGSG-----HLNPLILEELRVLDQRQI-HQMOMEOICROVLL	207	
DB	517 GSIVSSSIGGLSSRTTGPSSGLPSLSLESPLLAGITYSMKDPLISHSGGMRAVCSTPLS	576	
QY	208 LGSIGQTVGARASP-----SELPGTAGASTTKPLLPFPSPIKPAOTGKTITASSSSSS	259	
DB	577 SSLIGRP-GTSALPRLSRPFTSTLPSSSASISITTOVFSL-----AGSTFSL	622	
QY	260 SSSGAEPKQAFHLY-LPLGSOHPSPSVGVGSHKP-----TPAP	299	
DB	623 PST-----HTFGTPMAVNPLLSQAESSHTPEDLEDCSFRCRGTSPQESTLSSMP	672	
QY	300 SPALPGSTDOIASP-----HLAFPQT-----QLAAQCUGAANGLEA-	338	
DB	673 ISSIPALFDOTASACGGGOLDPRAPAGCTTMEOELLEOGDSEAGVNIYMKLKALLALOQE	732	


```

OY 339 -----AASPLLKPKNG-----SGELGYEVIS 361
DB 733 NORLQEQILSTAKKERLOILNVOLSVFPPALPALPANGPVPGRYGLPPQAGSSDLS 792
OY 362 SLEKPGGRHKCRFCAKVFGSDSALOILHRSHTGERPYK-CNVCGNREPTTGNLKVHFRH 420
DB 793 TSKSPGKSS-----LGLDNL-----STSEDPHSGC-----PSRSSSLSFHST 833
OY 421 REKYPHVOMNHPVPEHLVDYITSSGLPYGMSVPPKAEFEAGTGGGVKRPVLVASTTA 480
DB 834 PPLPLLOQSPATLP-----LALPGAP--APLPQ-----PONGIGRAAGAGLGA 877
OY 481 LSATESLT--LSTGTSTAVAPGLPTFNKFLMKAVEPKSKADENTPPSGSAI---AG 535
DB 878 MPMAEGLLGLAGSG-----GLPL-----NGLLGLNGAANPNPAS 913
OY 536 VADSGSATRQOLSKLVTSPLSMALLTNHLKSTGSPFPYVLEPLGASPSSTKIQ---Q 591
DB 914 LSGAGGAPTLQLPCLNLSLTE---QQRHLLQOQEOQLQO-LQQLLASPOLTPHQVTVYQ 969
OY 592 LVEKIDROGAVAVASTASGA--PTTSAPAPSS---SASGPNOCVLCRLVLSCPRLRLH 645
DB 970 MIOQIOQKRELRLQMAAGSOLPMASLLAGSSTPLLSAGTPG---LLPTASAPPL--- 1022
OY 646 YGQHGERPFCKVCGRAFSTRGNLRAHFVGHKTSPPARAONSCPICQKFTNAVTLQOH 705
DB 1023 -----PAGALVAPSLGNNTSLMAAAAAAANAAGGPPVLTQTIN 1062
OY 706 VRMHLGGQI-----PNGSALSEGGAQOE 730
DB 1063 PFLSLGAEAGSGGPKGCTA-DKGSANQOE 1091

RESULT 37
PCT-US94-04496-55
Sequence 55, Application PC/TUS9404496
GENERAL INFORMATION:
APPLICANT: Croce, Carlo
APPLICANT: Canaan, Eli
TITLE OF INVENTION: diagnostics, therapeutics and methods
TITLE OF INVENTION: for detection and treatment of acute leukemias
TITLE OF INVENTION: Resulting from Chromosome Abnormalities in the ALL-1
NUMBER OF SEQUENCES: 86
CORRESPONDENCE ADDRESS:
ADDRESSER: Woodcock, Washburn, Kurtz, Mackiewicz &
ADDRESSER: Norris
STREET: One Liberty Place, 46th floor
CITY: Philadelphia
STATE: Pennsylvania
COUNTRY: USA
ZIP: 19103
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/04496
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Deluca Esq., Mark
REGISTRATION NUMBER: 33,229
REFERENCE/DOCKET NUMBER: TJU-1242
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 568-3100
TELEFAX: (215) 568-3439
INFORMATION FOR SEQ ID NO: 55:
SEQUENCE CHARACTERISTICS:
LENGTH: 1093 amino acids
TYPE: amino acid
STRANDEDNESS: single

```

```

; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ;
; HYPOTHETICAL: NO
; ;
; ANTI-SENSE: NO
; ;
PCT-US94-04496-55

Query Match 4.0%; Score 211.5; DB 5; Length 1093;
Best Local Similarity 20.6%; Pred. No. 4,1e-06;
Matches 192; Conservative 88; Mismatches 321; Indels 329; Gaps 41;

OY 3 QETGSSSRILGPGCEPAREGCDASEHNHPQVACCAQAFSDPTEFLAHONSCCTDPYMW 62
DB 289 QETSESSR-----ESKSKSSSHLSLHKCKTLSSGCGVSFTTASSSSSSSSS- 338
OY 63 IIGQENPNSNSASAPPEHSHR-SQVMDTEHSNPPDSSGSPDPPT----- 109
DB 339 -GGPQPAVSSLOSPPFSAPFKLEQPEEDYKSTPAAPASAPPSAPDEPPKADLFEQ 396
OY 110 -----WGPERR---GESSSQFLVAATG-----TAAGGGGLILASRLGATPLPE 153
DB 397 KYVFSGFGPIKRFSTTTSSGRARAPSPEDYKSPHYTSGASAGTKRMALSATPVAD 456
OY 154 STP-----APPP-----PPPPPPG----- 169
DB 457 ETPETGLKEKKHKASKRSRHGPRKSGSRNKKGTTGGPAPAPSLPSAQLAGTATAPSPSG 516
OY 170 ---VSGS-----HLNPLLEELRYLQQRQI-HOMQTEQICROYL 207
DB 517 GSLVSSGLGLSRTFGPSGLPSLSLESPLLGAGIYTSNKPDIISHGGMIRAVCSPLS 576
OY 208 LGSIGQTVGAPASP-----SELPGTAASSTKPLPLFSPIKRAQCKTTAASSSSS 259
DB 577 SSLGPP-GTSALPRLSKSPPTSLPSSASISTQVESL-----AGSTFSL 622
OY 260 SSSGAEPPKQAFHLY-HPLGSHPFVSQVGRSHK-----TPAP 299
DB 623 PST-----HIFGPMGAVNPLLSQAESSHTEPDEDCSFRORGTSPQSLSSMSP 672
OY 300 SPALPGSTDOLASP-----HLAPFGTT-----GLLAQCCLGAARGLA- 338
DB 673 ISSLPALFDQTASPCGGQOLDPAPAGTTNMEOLLEKQDGEAGVNIIVELKALHLOKE 732
OY 339 -----AASPLLKPKNG-----SGELGYEVIS 361
DB 733 NORLQEQILSTAKKERLOILNVOLSVFPPALPALPANGPVPGRYGLPPQAGSSDLS 792
OY 362 SLEKPGGRHKCRFCAKVFGSDSALOILHRSHTGERPYK-CNVCGNREPTTGNLKVHFRH 420
DB 793 TSKSPGKSS-----LGLDNL-----STSEDPHSGC-----PSRSSSLSFHST 833
OY 421 REKYPHVOMNHPVPEHLVDYITSSGLPYGMSVPPKAEFEAGTGGGVKRPVLVASTTA 480
DB 834 PPLPLLOQSPATLP-----LALPGAP--APLPQ-----PONGIGRAAGAGLGA 877
OY 481 LSATESLT--LSTGTSTAVAPGLPTFNKFLMKAVEPKSKADENTPPSGSAI---AG 535
DB 878 MPMAEGLLGLAGSG-----GLPL-----NGLLGLNGAANPNPAS 913
OY 536 VADSGSATRQOLSKLVTSPLSMALLTNHLKSTGSPFPYVLEPLGASPSSTKIQ---Q 591
DB 914 LSGAGGAPTLQLPCLNLSLTE---QQRHLLQOQEOQLQO-LQQLLASPOLTPHQVTVYQ 969
OY 592 LVEKIDROGAVAVASTASGA--PTTSAPAPSS---SASGPNOCVLCRLVLSCPRLRLH 645
DB 970 MIOQIOQKRELRLQMAAGSOLPMASLLAGSSTPLLSAGTPG---LLPTASAPPL--- 1022
OY 646 YGQHGERPFCKVCGRAFSTRGNLRAHFVGHKTSPPARAONSCPICQKFTNAVTLQOH 705
DB 1023 -----PAGALVAPSLGNNTSLMAAAAAAANAAGGPPVLTQTIN 1062
OY 706 VRMHLGGQI-----PNGSALSEGGAQOE 730
DB 1063 PFLSLGAEAGSGGPKGCTA-DKGSANQOE 1091

```



```
RESULT 38
US-07-945-283-2
: Sequence 2, Application US/07945283
: Patent No. 5352596
: GENERAL INFORMATION:
: APPLICANT: Cheung, Andrew K.
: APPLICANT: Mesley, Ronald D.
: TITLE OF INVENTION: Pseudorabies Virus Deletion Mutants
: TITLE OF INVENTION: Involving The EP0 and LIT Genes
: NUMBER OF SEQUENCES: 7
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Curtis P. Ribando
: STREET: 1815 No. 5352596th University Street
: CITY: Peoria
: STATE: IL
: COUNTRY: USA
: ZIP: 61604
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/07/945,283
: FILING DATE: 19920911
: CLASSIFICATION: 424
: ATTORNEY/AGENT INFORMATION:
: NAME: Ribando, Curtis P
: REGISTRATION NUMBER: 27976
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 309-685-4011 ext.513
: TELEFAX: 309-685-4128
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1958 amino acids
: TYPE: AMINO ACID
: TOPOLOGY: 1linear
: MOLECULE TYPE: protein
US-07-945-283-2

Query Match      4.08; Score 211.5; DB 1; Length 1958;
Best Local Similarity 20.2%; Pred. No. 8.7e-06;
Matches 196; Conservative 75; Mismatches 362; Indels 339; Gaps 43;

OY      5  TGSSRLGSGCGPRAEGRGDASEEN-----PQVCAKCAQPSDP-----TEFLAH 50
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      311  SGPSTSHQDRPRGPTP---STSHNNHNNQGRPTSPRSTSSHQDRPGGRRPASETHNNH 367
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY      51  QNSCSDPRVYVITIG-----QENPSNSASSAPRRECHSRSGVMDTEHSNPD 99
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      368  Q-----DPR-----GGRRPTSSHHNNHDDPRGGSPRRPRRSTSSSS-----SHGGRPS 412
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY      100  SGSSGP--PRDTWGR-----BRGSESSGQFLVAMT-----GTAA 132
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      413  TRRRPRQRPRRRPRRRPRRQKISTETRASGEENTAOULFSHSEKLFHNPJEGEGEGDGTAG 472
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY      133  GGGGGLILSPKLGATRLPRESTRAPRRRRPR--PRPGVSGHLNPLILELRYLQQRQ 191
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      473  GEGDRDPRPR--SPRRPRRLRLRRRRRRRRRRRRPRRPGASAR-----RRRR 517
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY      192  IHOMQTEQICRQVLLLSLQTYGASPSSELTGTGA-----ASTKPLRLFSF 242
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      518  -----GGGGGP----PGRGRRRGRGKRRRRRAAGTGAADAEE 550
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY      243  IKPAQTGKTASSSSSSSSGAEPRKQAFHLLYNPLGSONHPSVGGVGSRSHKPTPAF--SP 301
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      551  EEDGDEDEDDRAEDDEGREGGEGRRGA-----GGGAGESESESESSRAE 595
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY      302  ALPGSTDDQLASPHLAPRTTGLLAACSLGARGLAASAPSLKPKNGSGELGTEVYS 361
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      596  GABRSASEO-----OVGAVAGYLLLVNRGILLDGPRAAGAVAAAE--ADDLHRGRVLP 648
```

```
OY      362  SLEKPGR-----HKRCRAKVFSGDSALOILHRSHTGEPKYKCNVCGNRFTTRGNKV 415
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      649  VLAGPPGARGPYGLHG-----AAGGADAGLEGKRVANGRRGARGRGANGORGLGV 704
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY      416  HFHRRH-----EKYPHVNPH----- 432
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      705  GLOQRGAGGAGLQROALGAELTGEPARAAGDEDFGORGAEPPAVGAVREGARVKRV 764
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY      433  FVPEH--LDYIYS--SGLPYGMSPPEKAEFEAGTPGGVEKRLPYASTALATESLT 488
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      765  PEPERGALAGHVAHVGGGHALVAGARRORDRGPBGAGAH--VAHVYLAEAQRK- 820
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY      489  LLSTGTAVVAPGLPTFNKFLMKAVEPKSKADENTPPGSEGSALAGVADSGSATRQLS 548
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      821  -----GPGV-----QAGE-----GGLHGEAGRAHDGAVVDGGR 850
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY      549  KLVYSLPSMALLTTHLKSTGSR-----PPRYVLEPL-----GA 581
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      851  ELAALGPAHAGALGGRVQADVDVVPHGRAVRGPLYLDGVQHDPPARRAEPRAEVLDA 910
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY      582  SPSETSKLOQLEKIDRQGAVALVASTASGAPTTSAAPSSSSASGPNOCVLCRLVLSGPA 641
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      911  GEAEVPRREQ-----QHPLGEADVAP--GPVP-----GPG-----VRV 944
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY      642  LRLHYGNGER---PFCKVCYGRAPSTRGNLRAH--FVGHKTSPAARAQNSCPIQCKFT 697
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      945  RRAEAVGEGGEGQREBAAARVPCRGALGLGAEELLVGQR----- 985
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY      698  NAVTLOQHVYMHIG--GQIPNGSALSSEGGAAOENSEOSTASGSGFPPOSOQSPSEE 756
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      986  ---VEHHNHANHLGVLRLPHPGAAABERG-----ANCPARAGVDROGGRVGE 1030
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY      757  EMSDEEDEDDEEDVTDEDSLAGRSGESGGEKAISYRGSESEVSCAEETATVAAPTT 816
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      1031  RRAPFGEEDLVHNEGAGHLGRAVGGEGRRGGRPRVGLAGR---DAAEAAGRGVLG--- 1083
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY      817  VKEMDSNEKAPO 828
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      1084  ---HGPERAPE 1091

RESULT 39
US-09-121-321-16
: Sequence 16, Application US/09121321
: Patent No. 6090783
: GENERAL INFORMATION:
: APPLICANT: Saiga, Akihiko
: APPLICANT: Orita, Satoshi
: APPLICANT: Igarashi, Hisanaga
: APPLICANT: Okumura, Kouichi
: APPLICANT: Sakaguchi, Gaku
: TITLE OF INVENTION: DNA MOLECULE RELATING TO SUPPRESSION OF
: TITLE OF INVENTION: GENE EXPRESSION AND NOVEL PROTEIN
: NUMBER OF SEQUENCES: 20
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: FISH & NEAVE
: STREET: 1251 Avenue of the Americas
: CITY: New York
: STATE: New York
: COUNTRY: USA
: ZIP: 10020
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/121,321
: FILING DATE:
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/933,803
```


Db 500 CPGGLRSLLLHGARRSKPYGCEGCKSGVARRKSLIIHRSHTKERPYECAECEKSFNCBS 559
QY 412 NLKVHFRHREKYPHYVOMNPHVPEHLDYVITSSGLPYGMSVPEKAEFEAGTPGGVER 471
Db 560 GLRHQMTTHRGEPY----- 574
QY 472 KPLVASTTALSTELTLLSTGSTAVAPGLPTFNKFVLMKAVEPKSKADENTPPGSEGS 531
Db 575 ----- 574
QY 532 AINGVADSGSATRMQLSKLVTLSPSMALLTNHLKSTGSPPEPYLEPLGASPSSETSKLOQ 591
Db 575 ----- 574
QY 592 LVEKIDRGAVAVASTAGAPTTSAPAPSSASGPNOCVICTRLVLSCPRALRLHYGOHGG 651
Db 575 -----KCSECEKTYSRKEHLONHORLHTG 598
QY 652 ERPPKCKVCGRASTRGNLRAHFVGHKTSAPARAONSCPICQKKFTNAVTLQGHVBMHLG 711
Db 599 ERPFOCALCGKSFIRKQNL---LKHORIHTGERPYTCGCGKSFYRKESLKDHLRVHSG 654
QY 712 GOIPNGGSALSSEGGAQOENSSEQSTASGPGSFQPOQOQSPDEE 757
Db 655 G-----PGPGA---PROLPPPERD 671

Search completed: January 13, 2003, 15:23:49
Job time : 49.9776 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OW protein protein search, using sw model

Run on: January 13, 2003, 15:17:45 : Search time 22.4664 Seconds
(without alignments)
4287.595 Million cell updates/sec

Title: US-09-988-117-3
Perfect score: 5301
Sequence: 1 MAOETSSRLGCGCEPAE.....SSIPSGLSPPRKDDPTMP 1002

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 segs, 96134422 residues
Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database :
1: p1r1: *
2: p1r2: *
3: p1r3: *
4: p1r4: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1282	24.2	1323	2	T30253 spat protein - mo
2	1265	23.9	1350	2	T30341 zinc finger protei
3	1142.5	21.6	1061	2	JC7116 Xsal-3 protein - A
4	759.5	14.3	1355	2	S40022 spat protein - fr
5	742.5	14.0	1263	2	T13805 spat-related prot
6	712	13.4	1402	2	S42748 finger protein - f
7	377.5	7.1	744	2	T20969 DNA-binding prote
8	349	6.6	2282	2	T42717 prockir2 - chicken
9	330	6.2	1173	2	S70620 finger protein bow
10	322.5	6.1	744	2	S70619 transcription regu
11	317.5	6.0	1615	2	JC6510 ras-responsive ele
12	316	6.0	1727	2	T42717 gene handsight pro
13	311.5	5.9	1920	2	T13893 hypothetical prote
14	307.5	5.8	1891	2	T13594 finger protein ZNF
15	306.5	5.8	654	2	A5785 hypothetical prote
16	300.5	5.7	644	2	T46277 hypothetical prote
17	300.5	5.7	2232	2	T34434 DNA-binding protei
18	292	5.5	624	2	S41688 atrophin-1 - human
19	292	5.5	1184	2	G01763 finger protein 2,
20	290	5.5	651	2	B32891 finger protein - A
21	290	5.5	1350	2	S00647 hunchback-related
22	287.5	5.4	982	2	T43676 C-terminal domain
23	286.5	5.4	1173	2	T31421 atrophin-1 - human
24	283.5	5.3	2688	2	S50832 zinc finger protei
25	281.5	5.3	1184	2	S50832 C-terminal domain
26	281	5.3	1191	2	S53305 atrophin-1 - human
27	279.5	5.3	671	2	JE03288 zinc finger-type zinc
28	278	5.2	3942	2	T42730 Bassoon protein -
29	277	5.2	2715	2	T13049 eyeld - fruit fly

30	271	5.1	1060	2	S33641 homeotic protein z
31	269	5.1	1042	2	A31591 transcription regu
32	265.5	5.0	615	2	S05546 finger protein (cl
33	265	5.0	803	2	S26823 zinc finger protei
34	263.5	5.0	686	2	A34612 zinc finger protei
35	263	5.0	856	2	A53503 B-lymphocyte-induc
36	262.5	5.0	710	2	I48668 zinc finger protei
37	262	4.9	604	2	S05447 finger protein gla
38	261.5	4.9	536	2	S06548 finger protein (cl
39	261.5	4.9	675	2	S51037 zinc-finger protei
40	261.5	4.9	2500	1	WMHDE2 HIV-EP2 enhancer-b
41	261	4.9	428	2	S03677 finger protein (cl
42	261	4.9	581	2	A49073 HSN motor neurons
43	259.5	4.9	789	2	A35964 transcription repr
44	259.5	4.9	794	2	S59069 Z13 protein - mous
45	257.5	4.9	594	2	JC5146 arylphorin gene-sp
46	257.5	4.9	3938	2	T42761 Bassoon protein -
47	257	4.8	2529	2	A56923 transcription fact
48	256.5	4.8	485	2	A40751 finger protein MZF
49	255	4.8	553	2	S22954 finger protein zfp
50	255	4.8	908	2	T16057 hypothetical prote
51	254.5	4.8	595	2	G02075 transcription repr
52	254	4.8	439	2	S06556 finger protein (cl
53	254	4.8	591	2	S65088 finger protein XPO
54	253	4.8	494	2	A42170 zinc finger protei
55	253	4.8	543	2	E88280 protein egl-43 lim
56	250	4.7	411	2	S10245 finger protein, te
57	250	4.7	2578	2	A56922 transcription fact
58	249.5	4.7	1706	2	I84499 zinc finger protei
59	248	4.7	595	2	JC7779 Kruppel-associate
60	246.5	4.7	1051	2	A60191 oncogene Evi-1 - h
61	245.5	4.6	427	2	I49603 transcription regu
62	245.5	4.6	459	2	I38600 zinc finger protei
63	245	4.6	477	2	A47236 zinc-finger protei
64	245	4.6	1042	2	S41705 Evi1 protein - hum
65	244	4.6	542	2	A54661 zinc finger protei
66	244	4.6	614	2	JH0500 zinc finger protei
67	244	4.6	2717	2	A34203 DNA-binding protei
68	243.5	4.6	457	2	JC5076 myc-associated zin
69	243	4.6	582	2	S08666 finger protein ZFP
70	242	4.6	488	2	S47072 finger protein HZF
71	241.5	4.6	399	2	S47071 finger protein HZF
72	241.5	4.6	3190	2	CREB-binding prote
73	241	4.5	693	2	I37570 zinc finger protei
74	240	4.5	457	2	C57785 zinc finger protei
75	240	4.5	701	2	T14757 hypothetical prote
76	239	4.5	509	2	JH0501 zinc finger protei
77	239	4.5	636	2	I48689 gene NK10 protein
78	238.5	4.5	466	2	T08674 probable finger pr
79	238.5	4.5	3869	2	A48205 ALL-1 protein +GTE
80	237.5	4.5	625	2	D87793 protein C27A12.2
81	236.5	4.5	480	2	JC7812 BCL6 homologous z1
82	236.5	4.5	688	2	A56360 zinc finger protei
83	236.5	4.5	2783	1	A41948 alpha-fetoprotein
84	236	4.5	1721	1	I38598 retinoblastoma bin
85	234	4.4	589	2	I38598 zinc finger protei
86	233.5	4.4	540	2	B57785 zinc finger protei
87	233.5	4.4	546	2	I49636 DNA-binding protei
88	233.5	4.4	634	2	T47156 hypothetical prote
89	232.5	4.4	420	2	S06579 finger protein (cl
90	232	4.4	1117	2	JC4934 delta-crystallin/E
91	231.5	4.4	1634	2	T26517 hypothetical prote
92	230.5	4.3	753	2	S48059 metal-regulatory t
93	230.5	4.3	1032	2	T34433 hypothetical prote
94	229	4.3	572	2	I39311 Kruppel-type zinc
95	229	4.3	899	2	B48586 suppressor of hair
96	229	4.3	1151	2	T18535 high molecular mas
97	228.5	4.3	728	2	A54603 transcription fact
98	228.5	4.3	1585	2	T31611 hypothetical prote
99	228	4.3	577	2	S72227 finger protein sob
100	228	4.3	3968	2	A44265 trithorax homolog

QY	199	EOICHOVLLLG----	SLGOTVGAASP-----	SELPGTGAASST	233		
Db	247	EOIRQOVALMNOPLRLPLNPIVPSONAPRLPASNQLOGFAHSTLDLTSTVPPRLSGPAT			306		
QY	234	KPLRLPLFS-----	PIKPAOTGKTATSSSSSSSSSGAEP--	KQAFPH	273		
QY	307	SGLPSPFENQHMSPGSSGAFENICPVASVPTFESTISLTSTNKAASSAAPSLSLSTNSN			366		
QY	274	LHNLPGSQHPFVGCGVGRHKHRTPARSPALPGSTDOIILASPHLAFPECTGGLAACGAA			333		
Db	367	PTHPOSSSTPRLPGHGNILNSSSLSPRLPOS-----	SSNSVTFPPPLASTIAA-----	TA	417		
QY	334	RGLEAASPGLKLPKNGS--	GELGCGEVIISLEKPGGRHKRCFAKVGSPDALGHLRSH		392		
Db	418	NALDPLSA--	LKMHKRGKPRNVSEFTKTTSDDRPFENKRCFAKVGSDALGHLRSH		475		
QY	393	TGERYYKCAVCGNRPFTTGNLKVHFRHREKRYHVNOMPHRPVREHLDYVITSSGLPYGMS			452		
Db	476	TGERPFKCIKCNRESTKGNLKVHFRHREKRYHVNOMPHRPVREHLNGPTSSGIPYMS			535		
QY	453	VPRPEAEBAETPGCGVERKPLVASTATASATSLTSLTGTSTAVANPGLPTFNKFLYLMK			512		
Db	536	LPRERPVYT-----	WLDKRPVLPY--	VTITGLDLPY-----	IPGMGVASYSDSP	580	
QY	513	AEERKSKADEN--	TPPGSEGAIA-----	GVADSGS	541		
Db	581	STPNRSRPFQRPASSECSNLSPTNINSELCIQAASESPQEQRTPTVTKQERTVQSS			640		
QY	542	ASR-----	MOLSKVLT--	SLRSMALLTNILKST--	GSFPPRYLPLGA	581	
Db	641	SERVEQRYNVQIISSEVTTVPPLVTDSVSTSHSNVSLPRMSDFAKKFPFGGLEEM--			698		
QY	582	SLETSKILQOLLEKIDROGAVAVASTASGAPRTTSABAPSSASAGPNOCVLCRYLSCGRA			641		
Db	699	QSSSESKLQOLLENIIDKK-----	MTDPNCVICHRLSCHSA		735		
QY	642	LRLHNGONGEERPFCKYCGRAFTGRGNLRAHNVGHKTPSAARAQNSCPICQKKFTYNAV			701		
Db	736	LKMHRTTGEERPFCKYCGRAFTTGNLKTNGVHRKSRPLRLVQHSRPCIOKKFTYNAV			795		
QY	702	LOOHVRLMLGQIIPNGSALISGGGAQO-----	ENSESGS---		736		
Db	796	LQOHTIRMHMGQIIPN--	PLRPGFQNAKSELSYDDKNLETMSNYDDDFDONSLEDDDL		853		
QY	737	-----	TASGGSFPPOSOQPSPEENSEEEE-----	EDEEE	769		
Db	854	KOTASDSKRLPYGSSSPASSPTVISTIAALENQMIMDSVTAQOFIghkNIENGSE			913		
QY	770	EDVPEDESLAGRG---	SEEGGKKAITSVRDSEEVSGAEEEAVALSVAPRTVYKEDNSDEKA		826		
Db	914	IDHLSNDSSSAVGDLSEOSAGSPANSESSSMOVLSPRANSHSSTIRKSPV--	ISSQEP		971		
QY	827	POHTLPPRRPRLNDHPROMQGTSDVSGAMEBEAKLEGISSPMALPTQEGESTPRLV			886		
Db	972	PVIOIKTEKP-----	DSPIR--	PRENDGYLD-----	LTGTNPG--	RPII	1006
QY	887	EEELNPEA-----	AKKDPGESSGRKACVCGSPFTQALAEHQHTPRDGLFTVCFC		940		
Db	1007	KE---	EAPYSLFLSREGRKFS--	TVCNICGPFACSKALETHYRSHTEKRP--	FICTVC		1060
QY	941	ROGLFDRATLTLLKNNMLAHNQV-----	PPR-----				964
Db	1061	KRGCGTMONIKOHLLT--	HKLELPSQLFEPRNPTLEPQSQTTSVLVSTAVMIKMEVNGH				1118
QY	965	-----	APHGQNIATLSLVPGSSSITPSGSLSPFRK		996		
Db	1119	TKPISLAGEPHLPAGIOVLAAPOTAMSPGITEPMLAPPRR			1158		

[illegible]


```

b 6/4 -----SKPTPHSETRAESPAMGOSTGGOQKPVTLPLSALNLQROKSVKSSNGSLE 725
y 790 AIVSKGDSSEVYSAAEVATSVAPT-----TVKEMDSNEKARQHTLPPRPDPDN 840
b 726 SDGLTNDSSVWDOEYPTGKS---PTOSEANTFSPNTSOSDSNAS-----KSPSYNG 775
y 841 LD-----HPQ-----PMEOGTSDVS-----GAM 858
b 776 LDDLGLMSKDEHSQNSLNPDDGDLALDLTLNGGFAKKIKEEPGLHONGEFGKLEPNLYVGAP 835
y 859 EBEAKLEGISPYMAALTOE-GEHSTPLVEELINPEAKMKDGESSGRACVEGOSGFPT 917
b 836 PALIMKEVSDRAGATQYLGRPLMSGLNPLNLYQOR-----RSKQKHNTCTMGKKNFSS 889
y 918 OTALEBOKHTHPKDGPLTVCVFCROGLFDRATLKKHM 954
b 890 ASALQIHERHTHTGEKP-FACTTCGRALFTTKNKLKHY 925

```

RESULT 4
40022

palt protein fruit fly (*Drosophila melanogaster*)
 ;Species: *Drosophila melanogaster*
 ;Date: 19-Jul-1996 #sequence_revision 26-Jul-1996 #text_change 21-Jul-2000

Accession: S40022
 Author: Kuehnlein, R. P.; Frommer, G.; Friedrich, M.; Gonzalez-Galtan, M.; Weber, MBO J. 13, 168-179, 1994
 Title: spalt encodes an evolutionarily conserved zinc finger protein of the spalt gene
 Reference number: S40022
 MID:94139659, PMID:7905822
 Molecule type: DNA

```

;Residues: 1-1355 <KOE>
;Cross-references: EMBL:X75541; NID:9414106; PDB:CAA53229.1; PID:92568394
;Note: the authors translated the codon GAC for residue 51 as Ala
;Note: mRNA sequencing has also been done
;Genetics:
;Gene: sal; spalt
;Cross-references: FlyBase:FBgn0004579
;Introns: 51/1; 1329/3 1355/2
;Keywords: zinc finger

```

Query Match	14.38;	Score 759.5;	DB 2;	Length 1355;
Best Local Similarity	22.78;	Pred. No. 1.2e-28;		
Matches 312;	Conservative 141;	Mismatches 455;	Indels 505;	Gaps

3 QETGSSRLGGPCGPAERGGDASEHHPOVCAKCAQFSDPTFLAHONSCCTDPPVMV 62

50 KDIGSDQEEENGCCSPLTTATTTTASPSRPE-----PEEEQEEQSTSDSI---PEQST 100
63 IIGQENPSNNSASSAPRPEGHRSQVMDTEHSN---PPDSGSSGPPDPTWGPERRGEES 119

```

101 PDHOLENDIKSEAKSEIEPVEDNNNRVAMTKPSSSEEREPNASGSMSPSY--AEASAEAA 158
      |||  :  ::  ||  :  ::  ||  :  ::  ||  :
100 ECAETVATCTCAACCGCGCTTACDRTCAED-----LDRCEMDADDDDDDDDD 166
      |||  :  ::  ||  :  ::  ||  :  ::  ||  :

```

```

120  SGGFLVAAIGIAAGGGGGLLASFNLSAIF  100
      ::  ||  ::||::  |
159  ATE-----RTPEKEKEKDVENVMPDEAASSAVPSTEVL  194

```

167 PPGVSGHLNIPLEELR-----VLQ 188
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195 PGAGA----PYLEAIQNOMAIAPAKTIANGSGADNEAMKQLAFLOOTLNQ 249
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976

```

189 OROIHOMTEOTICROVLL.....LGS LGTIVGA 217
|:::|:::|:::|:::|:::|:::|:::|:::|:::|

```

250 QQQLFQIQILQQLQSLALNQAKQEEDEETEDADQDEQDEQETDITYEEEEERLADMELRKA 309

218 PASPSELPGTGAASTKPLPLFSPIKPAQTGKTTASSSSSSSSSSGAEPKQAFHFLYHP 277

```

310  EARM-----AKARQHLINAGVPLR-----ESSGSPAESLKRREHHDH-- 348

```

278 LGSQHPSVGGVGSHKPTPA-----PSPALPGSTD--QLIASPH--LAFPGITG 323
 | : || | : || | | | |
349 --ESQPNRRTSLDNTHKADTAQDALAKLEKEMENTPLPFSGDLASSITTNNHDDLPEPNSLD 406

[illegible]

0Y 580 ---GASBSETSKLQOL-----VEKIDRQGA VAVASTAGAP 612

Db 695 IMHPQSSPGSOSHLDHLPTPGQLPREDEFFAERFPLNFTAKMLSPENHSPVRSRAGAL 754

QY	0.3	TTAAAT	-----	0.2
Db	755	RRGVRRPHHHHMMARSPFFNF	IKHEMAALLPRHSDNDSWENF	IEVSNTEYMKLKEI 814
QY	620	---SSASGNSOCVTCISRVLS	CPBALRYHNGONGERPRKCKVC	SGAASFGMLRAHFVGH 677
Db	815	MKNKTIISDNOCVCRVLRSCSKAL	DMHRIYRPHGSEPRPKCKICGA	FTTKGMLKTHMAVN 874

[illegible]

```

QY      763 EDE-----EED-----VIDESLAGR 761
      :|:      | |
Db      986 DYDDVSYSEHLSNSLLEQECDDRSRGDDFKSLLEQKLRI DATGVVNTNPVPRSSASSH 104

```

```

OY      782 GSESGEKA---ISVRGDSEEV---SGAEEVA-----TSVAAPPTVTKEMDS-----822
          |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
Db      1046 GHSVSTAPTPSPVHAASSQYIKRSSPARSEASQCALDITPRAAPTSSSSSSRSLPKK 1100

```

[illegible]

DB 1100 PVSFSLRPSFSSSHASANLISLSPFLVGLDCLPFGDGMHQQGQHLMDQQGZNNVNNH 1100
QY 844 -----PQPMEOGTSDVSCAMEEAKLEGISSNAL-----TQEG 878

```

Db 1166 AAAQHHHHQQMALDQHQEQLRREAELAQQKAAAAAAAAAAAAAQOTPRQARDQREQ 122
ov 879 -----EGSTSTPLVEELNLP-----EAMKK-----DPGESSGR 905

```

Db 1226 GPGAGPPNP¹LMGARPPFGMFPN²LEPPAT³QNMCMAMNO⁴IAQSYMPA⁵AFNPLALSGV 128

Qy 906 K--ACEVCGSPTQTALAEHQKTHKKDGPLFCVFCRGGLDRATLKHHML 955
: : : : : : : : : : : : : : : : : : :
Db 1286 RGSITGCICYKTFPCHSALEIHYRSHTKERP-FKCSICDREFTTGNLQHML 1337

RESULT 5

T13805
spalt-related protein - fruit fly (*Drosophila melanogaster*)
C:Species: *Drosophila melanogaster*
13-Nov-2000 #cont change 17-Nov-2000

C:Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #tex_change 17-Nov-2000
C:Accession: T13805
R:Barrio, R.; Shea, M.J.; Carulli, J.; Lipkow, K.; Gaul, U.; Frommer, G.;

Schuh, R.;

382 EPNSLDLOKRTQEVLD-----ASQGL-ANMADFA 414
356 YGEVIS---SLEKPGRRKRCACAVFGSDSALOILRSHTERPYKNCVGNRTTGN 412
415 FGDKSSDGKGNPEFFKRCRCYCGVFGSDSALOILRSHTERPYKNCVGSRTTGN 474
413 LKHFHNRREKYPVHOMPHRPHREHLDYITSSGLPYGMVPRPKAEKPGGVEK 472
475 LKVFHQAQFRPHVPMATPIREHMD---KHPRLDOQSPSSPTQSPATGL--- 527
473 PIVASTALSTESITLSTGTAVAGLPFEK--FVLKAVEPKSKADENTP----- 525
528 PRPSTSLTQOMQPMSEF---ASSPAFGLPGIYRPMELLSGATGSGTAGLHPRFP 583
526 --FGSGSALAGVAD-----SG----- 540
584 QMGGL-GAALKHHTDQSDMPDILKSSGSPSPHEEDNIAARLPVKSEMEKEHTM 642
541 -SATR-----MOLSKIVT----- 552
643 EAATRESAEERPLLEVRKIERIDEDQNLQEGMOKREPLTAVATPHPOCLIPTTAA 702
553 -SLP-----SWALTNLHLS--TGS----- 569
703 AKSPRSRLQCHARLSLWCSHPTTSMHACAVILGSGTHLDQLPTPDNVPRTPQREDFFA 762
570 --FPPY-----VLEPLG----- 580
763 ERPLNLTSTKTDHSPKSPGNAHNIERSPPFNPKHMAAFVPRHSDNSMENFTE 822
581 -ASPSSTKLOOLYEKIDROGAVAVASTAGARTTSAPADSSASPGMOCYICLVLSGP 639
823 VSTSTEMKLEKELK-----NKKISDPQCVCYCDKYLSC 857
640 RALRLYHONGEGERPKKVCGRASTRGNLRAHFVGHKTSPPARAONSCPIQKFTNA 699
858 SALOMHYRTHTGERPKRICGRAFTTKGNLKTMAVHKITPRPRNFQCPYCHKKYSNA 917
700 VTIQOHVAMHNG-----GOIPNG----- 717
918 LVLDQHLRLHTEPTDLTPEQIOAAEIRDRPRSMRGHFNPRFAAARHFGAMPGGAGG 977
718 -----GSLSEGG----- 725
978 PRGATGMPRGCHNGTGLSESSGDDLDNMDCGDDPDDISSEHLSNDRPATSDRBS 1037
726 -----GAQENSSEFQ--STASGPSF-----POPOSOOP----- 752
1038 DDEKSLLEOKLRIDPTGVNINSHORPHSAASNPNSIGSASAPSPAPTSPSSQRPSCS 1097
753 -----SPREEMSEEEDEDEEEDVTDE--DSLGRGSGSGGEKATSV----- 793
1098 PVRSSCSPPVRSVS---TSGQALDLTPRALPPRLASSSSSHSPYRQLLSVRRRLARSVS 1153
794 ----- 793
1154 SHRCVAVMVRALLSSQLRPSVIGIDCLRPGLQHHLQOQHNLMOQOXAATAAQAQHHNQ 1213
794 -----RGDSEEV--SGAESEVATSVAAPTVKEMDSNEKARQHTLPR--- 833
1214 MOQHAALHNOEHRLRBAEQVOOKAAEQOKAAQAAAAAAROROSPQ---PRPRSG 1270
834 -----PRPRPNLHROP-----MEOSTSVSGAMEEERAKLEGISPMALITOE 877
1271 ESSVGPAPRAPPLISARPPGMPRLPIRRPATQNMCMANMQIAQSWPARAPFPLIAS 1330
878 GEGTSPPLVEELNLPEAKMKDPEGSGRACAEVCGSGFPTQALAEHQKTHPKDGLPTC 937
1331 GYVGST-----TGICYKTFPCHSALEIHYRSHTERP-FKC 1366
938 VPCROGFLDRATLKKHML 955
1367 NICDRGFTTKGNLKHML 1384

RESULT 7
T20969
hypothetical protein fl5c11.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
R:Accession: T20969
R:Wilkinson, J.
submitted to the EMBL Data Library, April 1996
A:Reference number: 219352
A:Accession: T20969
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-744 <MIL>
A:Cross-References: EMBL:271260; PIDN:CAA95798.1; GSPDB:GN00019; CESP:fl5c11.1
A:Experimental source: clone fl5c11
C:Genetics:
A:Gene: CESP:fl5c11.1
A:Map position: 1
A:Intons: 26/1; 75/3; 108/3; 145/3; 251/1; 453/2; 500/2; 613/1; 649/1
Query Match 7.1%; Score 377.5; DB 2; Length 744;
Best Local Similarity 22.1%; Pred. NO. 8.1e-11;
Matches 182; Conservative 92; Mismatches 280; Indels 271; Gaps 29;
OY 22 GGDASEBHPQVCAKCAQESDPTFEFLAHQNSCTDPVWVITIGGQENPSSASSAPRP 81
DB 51 GGLPREDNRNLTPIHFVVFANPOFL---SLCAQ-----LGNSSRNVSSTAS--- 96
OY 82 ECHSRGVMPTHEHNPDPGSG---SGPDPWQPERRGEESGOFVAAATGTAAGGGGL 138
DB 97 -----TTSSCPLOSQOSQSFSSPALTHVLDAHEBOELFSDCVCTTSSNGDI 146
OY 139 -----ILASPKGATPLPESTPAPPPPPPPPPGVSGHNLPLLEELRVLOQR 191
DB 147 REHKCKOTLAS---RSTSVPSSTIPSSVCFSLPTTPCL-OFSTNESIGTSEIEDEDEE 202
OY 192 IHOMQTEQICROVLLGLSLQGTGVARPASPSLPGCAASSTRPLLFSTIKAQTKT 251
DB 203 DMDEGEHVAHQ--LEGHLLQ-----KS 224
OY 252 TASSSSSSSSGAPRKQAFPHLHPLGSHNPFVSGVGRSHKPTAPSPALPGSTPOLI 311
DB 225 DDKSKAMSLFNHAPRFAFPNMPPFLMRQRPD-----PRADVFA 265
OY 312 ASPHLAPFTGTTGLIAOCLGARGLEAASPGLLKPKNGSGELGYEVISLEKPGGRHK 371
DB 266 AGRHNDMDDEALMEISTDEAEKIRALY-----GDKAVPTDP---NQ 306
OY 372 CRCAKAVFGSDSALOILRSHTERPYKNCVGNRTTGRNLKVFHNRREKYPVHOMPH 431
DB 307 CILCRKRVLSCKSALOMHRTHTGERPKKICQRAFTTKGNLKTMHGVNRSKH----- 359
OY 432 HPVHEHLDYITSSGLPYGMVPRPKA-----EEEAGTPGGGY-- 469
DB 360 -----SFRGLP--ISLPRQLAAMHQHQIAPRQRIHINHPRTSAASAATAAQAQ 406
OY 470 -----ERKPLVASTALSTATE--SLT---LSTGSTAVA--PGLPTF----- 505
DB 407 IQASOOCRCIOQRFNLNGELAVHTEHNRSLTOPRVMPRTPTRVQTFPRVPRFTTTPS 466
OY 506 -----NKFLVMAKAVEPKSKADENTPRGSEGSALAGVADSGASTRQLSLVISLPRWA 558
DB 467 LNATDMSTQENLANILSAQLNDS--PMTDPSV-----EKKITRDPPMAASLSPS-- 517
OY 559 LILNHLKSTGSFPPRYVLEPILGASPSSTKLOOLVEK--IDROGAVA-----VAST 607
DB 518 ---NSDSSSS---VKODILSESEFEKLLKLEPRLIEQOVSTTPNKNENPLAMQ 569
OY 608 ASGAPTTSAAPSSASAGPNOCVICLRVLSCPRALRYHGOHGERPKKVCVGRASFSTR 667
DB 570 KMAAETEPFRPPQRPVLSKHQCGVCFKHFRSSSSSALOILHMTHTHGDKRPFKCMCGRAFTTR 629

668 GNRRAHFVGH--KTSPARA----- 685
 630 GNLVKMGHTSWOOSPSRRGRIRFDVASSYTEKPMGLSPILFTSGAPGASPLAMLGNGL 689
 686 -----QNSPCIOCKKFTNAVTLOOHVHMHGCGQIPNGSA 720
 690 SGLEMMMLMRYCVCQKQKVCSPNELFQHLKEHLN---NGSSA 730
 RESULT 8
 142717
 DNA-binding protein Rc - mouse
 N:Alternate names: Ig kappa chain gene enhancer Recognition component
 C:Species: Mus musculus (house mouse)
 C:Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 18-Feb-2000
 C:Accession: T42717
 R:Mu, L.C.; Liu, Y.; Strandtman, J.; Mak, C.H.; Lee, B.; Li, Z.; Yu, C.Y.
 Genomics 35, 415-424, 1996
 A:Title: The mouse DNA binding protein Rc for the kappa B motif of transcription and for
 ew family of large transcriptional proteins.
 A:Reference number: 222236; MUID:97001141; PMID:8812474
 A:Accession: T42717
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-2282 <MWL>
 A:Cross-references: EMBL:L46815; NID:g1377885; PID:g1377886; PIDN:AA040884.1
 A:Experimental source: strain BALB/c; clone T1; thymocyte, brain
 C:Genetics:
 A:Gene: Rc
 C:Function:
 A:Description: binds V(D)J recombination signal sequence and kappa B motif
 C:Superfamily: HIV-1p2 enhancer-binding protein
 C:Keywords: DNA recombination; transcription factor
 Query Match 6.6%; Score 349; DB 2; Length 2282;
 Best Local Similarity 23.1%; Pred. No. 6.2e-09;
 Matches 216; Conservative 100; Mismatches 316; Indels 304; Gaps 44;
 QY 203 ROYLLLGSGYTGAGASPSBELGTG---AASSTKPLPL--ESPDKPAQGTKTASSS 256
 18 KRLTKKGEALIQSVSSAP--YPSGTTAPSESATOLLAPFGSPSQEKTGQ----- 69
 QY 257 SSSSSSGAPRQAFPLHYPLGSHQHPFSGVGVGRSHKPTPA-----PSPALPGSTDOL 310
 Db 70 --QOKRARRPSLEASVHT-----SOLP-----QHPPLAFMSPGKPEHLLSGSTWOL 114
 QY 311 IASPLHAFPGTTGLAA-----QCLGAARGL---EAAASPGELKPKNGSGELCYGEV 359
 Db 115 VDPMR---PGPSGSFVAVPGSHQSQLPLRSHASILRPELRLPIKVFYRPSQVSLKPAEE 171
 QY 360 ISSLE---KPGRNHCRCRCAKYFGSDALQILHRSHTGEPRPYKCNVCGNRPFTGNLKV 415
 Db 172 AHKKRRKPKOP--GKYLCQYCSRPCAKRPSVLQKHRSNHEPRYPCGPGFSFKTSNL-- 228
 QY 416 HFHRRREKYPHYOMNPRVRENLDYVITSSGLPYGMS---VPPEK-----AEEB 461
 Db 229 -----YKHRSNHRIRKAGLASGSSSEMYPGLMEKRIIPGEFEFEPTGESTDSEB 280
 QY 462 AGTPCG-----GVEKRLVASTTALATE-----SLTLSTGTSTA-VAP----- 500
 Db 281 TGAAGSPSTDVLPKRNPLRSSLSYSGSHGSSQERCSLSQSSSTGSPSLDDPAFPAEASE 340
 QY 501 -----GLPFNKRVLVMKAYEPKSKADENT--PRSEGAALGVADSGATRNQLSKL 550
 Db 341 HPLSKRPEDTHTIKQALNLRLSERKRLIEOTELSPSGKSTEGGFGRSBSAEQOVSP 400
 QY 551 VTSLPSMALL-----TNHLKSTGSFPFVYLEPLA-----SPSESTKIO 590
 Db 401 NTNAKSYAEIIRIGKCGRIGQRTISMALASTQP---LLPLSSEDKPSLVLPSVPTQYIE 456
 QY 591 OLVEKID--RGCAVAVASTASGAPTTAPAPSSASGPNOC----- 629

```

Db 457 HIKRLINEAVNDTSEIDSKPRRSSILTRSSVSPKSSILXRDSLSSHGKTKQEOSSL 516
QY 630 -----VICRYLSCPRLRLHYGHONGBERPFCYKCGRAFSTGRLNRAHEVGNK 678
Db 517 SLOHPSSSTPRVPLRLRHSMPSA-----ACTISTHNHTFGSY--SPDDHV 560
QY 679 TSPARONSCPIQCKFTFNVTLOOHYRM--HLGGQIPNGSALSIEGGAAQZMSSEQ 735
Db 561 ADPEVSRKN- PV---FTS-----HRYMKRNHATELPLGG-----EYSSEE 598
QY 736 STASGPFSPPOPOQSPREEEMSEEEDEEEDVD-----EDSLAGRG 782
Db 599 -----PGSSKDPRT-----SKRDEPERKEDLTKTKKGFRTKCANECTICGAR 644
QY 783 -----SSSGEKAISVRCDSSEEVSGAEVY-----ATVSAAPTIV 817
Db 645 YKRRDNYEAHKYKYSLOITTKANSVGHNEYEKTOABEPMSOMNYKLGATLELPLRK 704
QY 818 KEMDSNEKAOHTLPRP---PRPDNDLPROMEGTSDVYSAMEEAKLEBISPMAL 874
Db 705 RR---KEKSLGDEEPRAPFACPGSSELAHNR-----LSTKSPAEL- 743
QY 875 TOEGESTSPRLVEE--LNLPEAMKDRQESSGRACEVCGOSFPOTALDEHQKTHPKKG 932
Db 744 -----SKSAPSLBEDPRASSPGLPQSOELGONGRG-EGCPKKFVYIGNHTSPREKSPRPQ 797
QY 933 PLFTCYFCRGFLDKRLTKKMLLHNOVPPRAPRG 968
Db 798 P-----SGLEEDKPRAPGSSPPRAPRG 819

```

```

RESULT 9
150620
prockr2 - chicken (fragment)
C:Species: Gallus gallus (chicken)
C:Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 21-Jul-2000
C:Accession: 150620
R:Schultz, B.; Miesing, J.
Gene 148, 227-236, 1994
A:Title: Cloning and structure of a chicken zinc finger cDNA: restricted expression 1
A:Reference number: 150620; MUID:95047430; PMID:7958949
A:Accession: 150620
A>Status: preliminary; translated from GR/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1173 <SCH>
A:Cross-references: EMBL:X56805; MID:g577018; PIDN:CAA40140.1; PID:g577019
A:Genetics:
A:Gene: ckr2

Query Match          6.2%; Score 330; DB 2; Length 1173;
Best Local Similarity 19.8%; Pred. No. 2.4e-08;
Matches 187; Conservative 81; Mismatches 278; Indels 398; Gaps 34;

QY 1 MAQETSSSRILGPGCEGPAERGGDASEEH-----POVCAACCAQFSPTEPLAIONSC 54
   | | | | | : : : : : | | | | | : : : : :
Db 58 LNHGTCERPKKCSCPKAFKNSSLRHRHHTHTGERPYVCSAGKAFYGTSTNLROHRT 117
   | | | | | : : : : : | | | | | : : : : :

QY 55 CTDPYVAVIIIGQENBPNSSASSAPRPEGHSRSQVMDTEHSNMPDSSGPPDPPTWGPER 114
   | | | | | : : : : : | | | | | : : : : :
Db 118 TGERPY-----ACSHCSKFTHTSSNLL-----HQ 142
   | | | | | : : : : : | | | | | : : : : :

QY 115 REEESGGFLVATGTAAAGGGGGLIASPKGATPLPPESTTAPPPPPPPPPVGVSGH 174
   | | | | | : : : : : | | | | | : : : : :
Db 143 RTHSSRSHKCPKCPKAEVSDACIQKHLQSHAASPL-----LPSPSPQSLSP----- 191
   | | | | | : : : : : | | | | | : : : : :

QY 175 INIPLLEELRVLQROIHOMOMTE-----QICROVLLIGLQGTVGAPSPSPETPGTAA 230
   | | | | | : : : : : | | | | | : : : : :
Db 192 ---PLLEAVEML-----YKCECTWLSRCQDEL--GHRSHRPQRPVGDV--TATA 237
   | | | | | : : : : : | | | | | : : : : :

QY 231 SSTKPLLPSPFKPAQTGKTATSSSSSS---SSSGAEP-----KQAFHLHPPLGSG 281
   | | | | | : : : : : | | | | | : : : : :
Db 238 THRCP-----TCGKTFKNSGLARHRHSHGAPRPKCSGCHRSFGQLAGLGH 286
   | | | | | : : : : : | | | | | : : : : :

```


282 HPFSVGVGRSHK-PTPAPSPALPGSTDOLIASPHLAPFGTTGLLAOCIGARGLEAA 340
b 287 -----RGSAETPPHPPTPT-----SVPS----- 308
341 SPGLLKPKNGSGELGYEIVSSLEKPGGRHKCRFCAKVFGSDALOIHLSHNGERYKC 400
b 309 -----ERP-----YCTEGKAPKSGSGLRHYMDHNGERYKC 342
y 401 NVCGNRFTTGRNLKVHFHRRREKYPHYQMNPHVPEHLDYVITSSGLPYGMSVPEKAE 460
b 343 SECPKAFKRKSSLLAIH-----QRVHTGLRAVKCP-----SCGLTFKWS----- 380
y 461 EAGTGGGVGRKPLVASTALSAATESLTLSTGTSTAVARGLPTFNKFLMKAVEPKSA 520
b 381 -----SHLQYHLRLHNGERYKCPDC-----KA 404
y 521 DENTPPSGSAGIAGVADSGSATRMQLSKLYTSLPSWALLTNHLKSTGSPFPYVLEPLG 580
b 405 FKNT-----SCLGRH----- 414
y 581 ASPSETSKLOOLEVEKIDROGAVAVASTAGAPTTAPADSSASGPNOCYICLVLSGPR 640
b 415 -----ROL-----HTGERPHACRPGICRAFTQTS 437
y 641 ALRLHYGNGGERPFPKCVGGRAFSTRGNLRAHFVGHKTPSPARAONSCPICOKKFTNAV 700
b 438 NLROHQRHTNGERYACSHCGKTFTHSSNLQHL-----QRTHSSAR-PRHOCPCRAFWAS 493
y 701 TLOOHVFMHLGGQIPNGSGALSEGGAOENSSSEOSTAGSGFPPOSOQPSPEENSE 760
b 494 YLQRHRLTH-----AAGPKGSPRP----- 512
y 761 EEEDEEEEDVYDEDSLACGSGSEGEKAISVRGDSSEVSGAEFEVATSVAAPTTVKEM 820
b 513 -----ALTRPRDGPVQLQAA-----LSLEYTAP----- 534
y 821 DSNE-----KAPQ-HTLPPRPPRPDNL-----DHPQMEQSTSDVSGAMEEAKLEGIS 868
b 535 DAHTFLLQTPPGQILPSPRAPQKLLPLTAPQRPKQSGSPPTGQSLLVSTGTT 594
y 869 SP---MAALTOEGEGTSTPLVEELNLPEAMKKDPGESSGKACE 909
b 595 LPTLRQAQVAVAPQGTGTGILVQLGRLPHRAGIRPHGQAAVE 638
RESULT 10
570619
finger protein bowel - fruit fly (Drosophila melanogaster)
Species: Drosophila melanogaster
Date: 19-Mar-1998 #sequence_revision 17-Apr-1998 #text_change 17-Nov-2000
Accession: S70619
R:Wang, L.; Coulter, D.F.
EMBO J. 15, 3182-3196, 1996
Title: bowel, an odd-skipped homolog, functions in the terminal pathway during Drosophila embryogenesis
Reference number: S70619; MUID:96272178; PMID:8670819
Accession: S70619
A:Molecule type: mRNA
A:Residues: 1-744 <MAN>
A:Cross-references: EMBL:U58282; NID:g1388165; PIDN:AAB17949.1; PID:g1388166
Genetics:
A:Gene: bowl
A:Map position: 2L
Keywords: transcription regulation; zinc finger
Query Match 6.1%; Score 322.5; DB 2; Length 744;
Best Local Similarity 18.3%; Pred. No. 3.3e-08;
Matches 195; Conservative 90; Mismatches 281; Indels 501; Gaps 37;
y 90 MPDHSNPPPGSSGPPPTWGPBERGESSGQFLVA-ATCTAAGGGGLIASPKLG-- 146
b 1 MPTESSESSEISGGGGAIPMLRPSR-----MDQFMNSMAAALAAVGGGLPGCAADRNGSS 55

QY 147 -----ATPLPESTR-----APPPPP 164
b 56 GGSDDGSONGNDSDRMSASRIATETOLAYQOHLAGLHGRPPPPPSHREISAFVPL 115
QY 165 ---PPPGVSGHLNIPLLEELRVLQOQIOMQMTQICQVILLGSLGQTVGARASP 221
b 116 PTGKVRPGSNSVEITIAMMADKREKALRE-----AAAAAAMIGRPGGCGP 163
QY 222 SELPTGAASSTKPLPLPESPIKPAQTGKTVASSSSSSSGAEPPKQAFHLYHPLGSQ 281
b 164 G-VPPPGV-----LGP-----AGVPP----- 180
QY 282 HPFSVGVGRSHKPTPAPSPALPGSTDOLIASPHLAPFGTTGLLAOCIGARGLEAAAS 341
b 181 -PYLIG-----GSPSTGAGS-----SP---FPPGAAAALFPPGLGPGMGHGLD 221
QY 342 PGLK- PKNGSGELGYEIVSSLEKPGGRHKCRFCAKVFGSDALOIHLSHNGERYKC 400
b 222 RLRLRAPGRAS-----RPKQFICKFCNROFTKSYNLIHERTHDERPYSC 268
QY 401 NVCGNRFTTGRNLKVHFHRRREKYPHYQMNPHVPEHLDYVITSSGLPYGMSVPEKAE 460
b 269 DICGRAFRRODHLRDH-----RTH----- 288
QY 461 EAGTGGGVGRKPLVASTALSAATESLTLSTGTSTAVARGLPTFNKFLMKAVEPKSA 520
b 289 ----- 288
QY 521 DENTPPSGSAGIAGVADSGSATRMQLSKLYTSLPSWALLTNHLKSTGSPFPYVLEPLG 580
b 289 ----- 288
QY 581 ASPSETSKLOOLEVEKIDROGAVAVASTAGAPTTAPADSSASGPNOCYICLVLSGPR 640
b 289 -----SEKPKFCIECGGFGQSR 307
QY 641 ALRLHYGNGGERPFPKCVGGRAFSTRGNLRAHFVGHKTPSPARAONSCPICOKKFTNAV 700
b 308 TLAVHKLHMEESPCHKPCVCSRSFNORSNLKTHLTH-----TDHKYECCSGGVFRNC 363
QY 701 TLOOHVFMHLGGQIPNGSGA-----LSEGGGA 728
b 364 DLRRHALTHAVGVNSGDYVDVGEDEARNLGDEEDSLLEVDSPROSPVHNLSGSGSG 423
QY 729 QENSSQSTASGSGFPPOQPSPEENSEEEDVDE-----EEDVYDE 776
b 424 EKSESRMRKRAIDHEES-----EEFDQDEEELQDLRVHDLPREDDDDFPRED 478
QY 777 -----SLAGRSES-----GGEK-AISVRGDSSEVSG 802
b 479 EQQAEVALVARFOASKAAATSSSSSVGTRPGQVTHCHNEGGEYTYMRPHGKHQEEP 538
QY 803 ABEVATSVAPPTTVKEMDSNEKAPQHTLPP-----PPR-----PPNDLHP----- 844
b 539 GNSGLASLVPSPFVR-----YSPVGAAGPPRAPRGARPPRPHQNHLPN 588
QY 845 -----QPMEOGTSVSGAMEEAKLEGI-----SSPMALTOEGEGTSTPLVEEL 889
b 589 GDPYLPLIHVRDLNHNLSINLSKA-----GVPRPPRTPPTITQPESG--KPLNDPL 638
QY 890 NLP-EAMKKDPGESSGKACACVCGSFPYOTALLEEOKHHPKQRPFLTYCFRCRGLDRA 948
b 639 HSPHEAMPSTFLGSIEMKRI-----LPAPLIDMDPHN--HPGLG-----QRTFVDS 684
QY 949 TLKKHMLAHNOVPPFAPHG--PONTATLSLVPGSSSIPSPGSLSPF 994
b 685 STIALNMSRH---PPROLGKRPSTETSGATTEKGPVAVAPRIAPP 728
RESULT 11
B60191
transcription regulatory protein Evi-1, short form - human
N:Alternate names: ecotropic viral integration site 1; oncogene Evi-1

C:Species: Homo sapiens (map)
C:Date: 03-Mar-1993 #sequence_revision 03-Mar-1993 #text_change 29-Aug-1997
C:Accession: B60191
R:Morishita, K.; Parganas, E.; Douglass, E.C.; Ihle, J.N.
Oncogene 5, 963-971, 1990
A:Title: Unique expression of the human Evi-1 gene in an endometrial carcinoma cell line
A:Reference number: A60191; MUID:90326419; PMID:2115646
A:Accession: B60191
A>Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-727 <MOR>
C:Genetics:
A:Gene: GDB:EV11
A:Cross-references: GDB:119889; OMIM:165215
A:Map position: 3q26-3q26
C:Keywords: alternative splicing; DNA binding; zinc finger

Query Match 6.0%; Score 317.5; DB 2; Length 727;
Best Local Similarity 25.7%; Pred. No. 5,5e-08;
Matches 140; Conservative 67; Mismatches 20; Indels 137; Gaps 29;

QY 367 GGRHRCFRCAKAFGSSALQIHLRS-HTGERPKYCNVCNFTTRGNLKVHFRHREKYP 425
Db 128 GKHYECENCAKAFYTPDSNLQHRHTRQHVGARAHACPECKTFATSSGKQHHHNSVYP 187
QY 426 HVQMP--HPRVEH-----LDYVITSSG---LRYGMSVP-----SKAEACGPGGG 468
Db 188 FIFSQSMYFFPDRDLRLSLPLKMEPOSFGEYKKLQKSSSEPFDLITRKDEKPLTP--- 244
QY 469 VERKPLVASTALSTATSESTLTLSTGTSTAVAPGLPTFNKFLMKAVEP-----KSKA 520
Db 245 VPSKPPV--TPATSDQPLD-LSMG-SRSRNSG-----TKLTPRKNHVFSGKGS 291
QY 521 DENTPPGSGSALAGVADSSAT-----RMQLSKVTLSPMALLTNLKSTGSGFP 572
Db 292 NVESRPASGS---LQHARPFFEMDPYRVEKRKLTLD--PLEALKRYLRPSPGLF 344
QY 573 PYVLPLPLGASPSF---TSKLOQLVEKIDROGAVAVASGAPTSAPASSASGPN-- 627
Db 345 ---HPQFOLPDRQTMMSALENNAEKLESFAL---KPEASLLOSVPMSFNPAPNAL 397
QY 628 -----QCVCILKLVLCSPRALRL--HYGQHGGERPKCKVCGAFSTGNLRAH 673
Db 398 PENLRLKRGKERYCARGCKGKIF--PRSANLTRHLRTHTEOPRCYCDRSFISSLQHRH 455
QY 674 F--VGHKTSPPARAQNSCPICOKKFTNAVTLQOHVBMGLG-----QIPNGS 719
Db 456 VRNHNHKEKPF---KCHLCYRCFGOOTNLDRHLKHKHNGMSGSTATSSPSELESTGA 510
QY 720 AL-----SEGGGAOENSSEOSTASPGSFQPOQSPSEEESEER 762
Db 511 LDDKEDAYFTEIRNFIGNSHGQSQPRNVEERNNGS---HFKERKALVPSQNSDLDDE 567
QY 763 EEDDEEEVDTDED-SLAGGSESSEGEKAIS-----VRGDEEYSGAEEVAATVAAAP 815
Db 568 EVEDDEVLDEDEDDYITCK--TGKEPVTSNLHEGNPEDDYEEETSALEMSCKTS--PV 621
QY 816 TYKE 819
Db 622 RYKE 625

RESULT 12
JC6510
ras-responsive element binding transcription factor RREB-1 homolog - chicken
C:Species: Gallus gallus (chicken)
C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 21-Jul-2000
C:Accession: JC6510
R:Miyaake, J.H.; Szeto, D.P.; Stumph, W.E.
Gene 202, 177-186, 1997
A:Title: Analysis of the structure and expression of the chicken gene encoding a homolog
A:Reference number: JC6510; MUID:98087435; PMID:9427563
A:Accession: JC6510

[illegible]


```

QY      618  APSSSASGPNOCVLCRLVSLCPRALRLHYGGGGERPFCKYCGRAFSRGRNLAHF--- 674
           || : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      1027  ---PPCGPVKMTIKNGVL-MPKOKQRRYRT--ERRPACSEHCASAPFTLASNMERHVKQ 1078
           || : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      675  -----VGKHTSPAARKQN----- 687
           || : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      1079  HPOFYAQRORSGHVHMRGRGASNVAAAAAATAVMAGPGSGSGFNHHGHGCHG 1138
           || : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      688  -----SCPICQ-----KFTNAVPTLQOHVR-----MHLGGQIPN--- 716
           || : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      1139  SHGSHGHAPRISDEQKCAILAOQLKAHKNTDLQALAHSSSVAGNPLLHFYPLTNPSP 1198
           || : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      717  --GGSALSBEGGA-----QENSSEOST-ASGPGSPPOPOSQOPSPEEEMS 759
           || : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      1199  MHNGS--SOGNGGATADIDDEPKLIIIDEDENHDEVEAEVDYDF--EEDDEDEMD 1251
           || : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      760  EEEEEE--DEEEELVTDSDS-----LAGRSGESGGE-----KAIVRG 795
           || : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      1252  EEEDEPELLIDEPFAEKEAEEOELPKPLEOLGTKEAQAQMAETLLEOAIKAKRPLSPPP 1311
           || : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      796  DSEEVSGAEEEVAT---SVAAPPT-----VKEMSNE-KAPQHTL 831
           || : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      1312  TENMAPANPTVATFTMOEPALITAPSTNPSSILKTMIAQAEYVGSKLKEVASSPPKDESODL 1371
           || : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      832  PPPPPPPNDHPQPMEGT---SDVSGAMEEAKLEGISSPMAALTOEGESTPPLVE 887
           || : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      1372  VPAKLVDNAT-SQNMGFNSYFRPSDVANHHMQSDS-EGIVASGSASESNSNGTEDVTSS 1429
           || : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      888  ELNLPEAMKKDPGESSRKACEVCGSFPETQTALEHOKTKRKDGLFTVCYFCRQGLDR 947
           || : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      1430  SSSSEPKKKSAYSILAPRKVSCPTCORMFPWSSILRRHIIITHGQKP-FKCSCHPLLETTK 1488
           || : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      948  ATLKKHMLLAHQV 961
           || : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      1489  SMCDRHLRLRHGNY 1502
           || : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 15
A57785
finger protein ZNF133 - human
C:Species: Homo sapiens (man)
C:Date: 23-Feb-1996 #sequence_revision 23-Feb-1996 #text_change 01-Dec-2000
C:Accession: A57785; S66506; S66509
R:Tommertup, N.; Vissing, H.
Genomics 27, 259-264, 1995
A:Title: Isolation and fine mapping of 16 novel human zinc finger-encoding cDNAs iden
A:Reference number: A57785; MUID:96044430; PMID:7557990
A:Accession: A57785
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-654 <TOM>
A:Cross-references: GB:U09366; NID:9487782; PIDN:AAC50260.1; PID:9487783
R:Vissing, H.; Meyer, W.K.H.; Aagaard, L.; Tommertup, N.; Thiesen, H.J.
FEBS Lett. 369, 153-157, 1995
A:Title: Repression of transcriptional activity by heterologous KRAB domains present
A:Reference number: S66506; MUID:95377390; PMID:7649249
A:Accession: S66506
A:Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-79 <VIS>
A:Cross-references: EMBL:U09366
C:Genetics:
A:Gene: GDB:ZNF133
A:Cross-references: GDB:137032
A:Map position: 20p11.22-20p11.23
C:Superfamily: zinc finger protein ZFP-36; LIM metal-binding repeat homology
C:Keywords: DNA binding; metal binding; transcription regulation; zinc finger
Query Match 5.8%; Score 306.5; DB 2; Length 654;
Best Local Similarity 20.2%; Pred. No. 1.6e-07;
Matches 172; Conservative 79; Mismatches 277; Indels 323; Gaps 32;

```


A:Map position: X	
A:Introns: 38/1; 75/3; 103/3; 132/2; 158/2; 222/1; 1088/1; 1367/1; 2039/1; 2049/1; 2075/1	
Query Match	5 7%; Score 300.5; DB 2; Length 2232;
Best Local Similarity	21.0%; Pred. No. 1.2e-06;
Matches 242; Conservative 133; Mismatches 453; Indels 327; Gaps 53;	
QY	6 GSSSRLLG--PCGEPAERG-----GDASEHHHPQCAKCAOCPDPEFLAHONCCCTDP 59
DB	390 GSSSYVGGSTTPASSSSSGTMSTMGSTGTVYAPVASSSTTGCSSTPIASSSSGCT--- 446
QY	60 VAVITG-----GOENPSNSA-----SSAPRECHSRNQVMDTEHSNPPDSC 101
DB	447 VTVVSGSSSTYGSTSPASSSSAGTASTIGSTGSTATIVPGSSSVSGSTQASPSSPG 506
QY	102 S-----SGP-----PDPWPERGEE-----SSQFLVAATGTAA 132
DB	507 TMTSTVSGPTGTVTVVPGSSTSPAPSSSPNPSSTPASTGSTITIGSSSIITVTVGSGTV 566
QY	133 GGGGGL---ILASPKLGATP-----LPPESTPAPPPPPPP-----PPGV 170
DB	567 SSGTGTSGSTLASST---ATPGSSSTVSSSPQSSQSPAPRTGTTSPQTSQSFSPSM 624
QY	171 -----GSHLNIPLILEELRVLQGRQIHOMQTEQICQVLLGLGQTVGAP---A 219
DB	625 NPSSSTPTGSS-----QSTTPEGSTASSPTGSTRFSVATEVT 664
QY	220 SPSELP-----GTCMASTPKPLPLFSPIKAOGKTASSS---SSSSSGAEP----- 266
DB	665 SOSTVPGSSGLGQSTNNS---PSPSLSPSTSGMSTLTPSPSSSTGSSQASTLTTP 720
QY	267 ---PQAFPHLYHPL-----GS-----QHPFSYGVGRSHKPPAPPALPGSTDLI 311
DB	721 SPNPQSTSLSESTSGATTSAGTMTSPSOSSSYGSSGCT---SPASTTGEMT 777
QY	312 ASPHLAPEGTTGLLAOCGLAANGLEAAAPG--LLEKRNKGELGYEYVSSLEKPGGR 369
DB	778 SGGSTGTGSSSVTSAAILTSTQOOSVSTNPGSTVTRPSTVSGSTSSGTVT----- 829
QY	370 HKCRCAKVFSGSALQIHLRSHTGERPKYKNCNRTTTRCNLKVHFRHKRYHYQM 429
DB	830 -----VGSTEA-----STGSSVA-----SSSPASTSQ 853
QY	430 NPHVPENHLYVITSSGLPYGMSVPEPEKAEER--GTPGGVERKPLVASTTALSTESLT 488
DB	854 NPNPSTSSGSMITQSPYPSQSTSPVSSSTTPSPGSPG-----TTLTSTSPSGSTT 906
QY	489 LLSGTSTAVAFGLPFENKFLVLMKAVEPKSKADENTPPGSEGSAT--AGVAD--SGSA 542
DB	907 IGTGSGT--SPGISTTSE-----EMTSOGSTQT--PGSTGTVTQPSVSDSTSSGT 956
QY	543 TRMOLSKLVTS--LPSMALTLNHLKSTGSPFPYVLEPLGASSETSLSKLOQVEKIDRGA 601
DB	957 VTVGSTEAGSSPLPSTSONTPSTSSS-----SMSTQPPQSSQSTSPV--ESS 1003
QY	602 VAVASTASGAP---TTSAPAPS--SSASGPNQVCYLRLVLSGPRALRLHYGOGGERPFK 656
DB	1004 TSGATSSSGSPGTTLNSIPSPSPSSSTIGSSQSTSPVYVSTISQSTERPSTGSGTVTKP 1063
QY	657 KVCYGRAFS-----TRGNLRAHFV-----GHKTSPARAONSCPICQKKFTNAVTL 702
DB	1064 STVSGSASGSTATMGSTEAASSTGGSSTSPNPSQSTSPSTGATSSPSSGTTLNSISP 1123
QY	703 QOHVNRHLG-----OIP--NGGSLSEGGAAGNASSQSTA 748
DB	1124 SPSSQSTISSQSTSPVYSTGDMTSQGTQLPGSTGTVTQPSSTGSGSTSGELTIS 1183
QY	739 SGPSPFPQPOSOQSPSEEMSEEEDEEEDVDYDDESLAGRGSESGEKAISVGDSE 798
DB	1184 QGSTQTPR--SLSTSPALISTGQOOSVSTNPGSTVTPQSTV--RGSTSSGSTVYTGTEGS 1241
QY	799 EYSGAHEEAVTSVAALPTTKEMDSNEKAPQHTLPPPPPPDNLHQPMEQGTSDVSGAM 858

DB	1242 STSGSSGATSLSSSPVPSTSQSPNPSTSGSSTPTPNP-----SQSTSPV----- 1286
QY	859 EEBAKLEGISPMALALQEGGCTSP--LVEELNLPPEAKKDRGESS-----GKKACEYCG 912
DB	1287 -----VSTTGEWMTSHG--STQTPSTIGSTVTPQSTVSGSNSGSGTVYIGSEASTSG 1337
QY	913 QSEPTQALAEHQKTHKRDGLFTFCVFCRQGLDRTATLKKHMLAHQVFPFAPHQONI 972
DB	1338 SSKFTSPS---SISPVTPSSPIPTTFASS--TSGSTIS-----DVSSV 1376
QY	973 ATLSLVPGCSSIPS 987
DB	1377 STTSLAP--LSSSLPS 1390
RESULT 18	
S41688	
DNA-binding protein - mouse	
C:Species: Mus musculus (house mouse)	
C:Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 05-Nov-1999	
C:Accession: S41688	
R:Fognani, C.; della Valle, G.; Babbiss, L.E.	
EMBO J. 12, 4985-4992, 1993	
A:Title: Repression of adenovirus E1A enhancer activity by a novel zinc finger-contai	
A:Reference number: S41688; MUID:94085372; PMID:8262041	
A:Accession: S41688	
A>Status: preliminary	
A:Molecule type: mRNA	
A:Residues: 1-624 <RG>	
A:Cross-references: EMBL:X76858; NID:9453618; PIDN:CAA54188.1; PID:9453619	
Query Match	5.5%; Score 292; DB 2; Length 624;
Best Local Similarity	23.2%; Pred. No. 7.5e-07;
Matches 141; Conservative 72; Mismatches 220; Indels 176; Gaps 26;	
QY	337 EAAASP-----GLLKPRNGSGELGYEYVSSLEKGRHKCFCAKVFSGDSALQIHLS 391
DB	3 EATGSPRHQELGLL-----GEGEQAHVKILVKNK-----GRVYCHLCHTFKGSILKAIMVT 55
QY	392 HNGERPCKVCGNCRFTTRGNLKVHFRHREKYPRV-----QMNPH-----PV 434
DB	56 HSSRKDHCKLCGASFRTKGLIRHHRHTDERPYKCAKGSFRSGALTRHLKSLTPC 115
QY	435 PEHLVDVITSSGLPYGMSVPEPEKAEERAGTGGGERKPLVASTTALSTESLTLSGT 494
DB	116 TEKIRPSTISDTAVAGKEEVPAGSSASIVGTYSVAQDMETSPV-----IHLVDAK 168
QY	495 STAVABGLPTFNKFLV--MKAVEPKSKADENTPPGSE-----GSAIAGVADS 539
DB	169 GYIHEVHVQMOELPLCMKALTTPESPOSEELPCSSENSRENLLHQAMONGIYLERVAGE 228
QY	540 GSAITRQOLSKLVTSLEPMALLTNHLKSTGSPFPYVLEPLGASPSSTSKLOQVEKIDRQ 599
DB	229 ESA-----LEPAPPS-----GSSP-----QCLGSGSPPLPLK--VEQIETQ 263
QY	600 GAVAVASTAGAPTPAPASSASGPNQV-----ICLRVYLSG-----PRA--- 641
DB	264 -----VASEAATVPPTH--PCPOCSETPPAATLLEAKRGTAPRPFTCTCGKAPPKAYLL 318
QY	642 -----LRLHYGOGGERPFCKYCGRAFSTRGNLRAHF 674
DB	319 KNDQEVHVERRRFCGCGCKLYKTITAHYGRHRRVHSDERFPQPCGGRKKTNAQOVHF 378
QY	675 VGHKTSPARAQNSCPLCQKKFTNAVTLQOHVNRHLG-----GOIPNG--GSAL 721
DB	379 RTH-----LEEKPHVCQCSRFREKGSILVRHVRHHTGKPRCYKCGRFPADGTLNRHL 434
QY	722 SEGGG-----AAQENSEQST--ASGPGSPFQPOSOQSPSEEMSEEEDEEED 770
DB	435 KTKGGCLLEVEELLVSESPSAATVLAEDHTVLVQSSVYADVQVYITTEATADTETS 494
QY	771 DVTDEDSLARGSE-----SGGEKAI--SVRGDSEEVSGAEEVA--TS 810

[illegible]

```
QY 564 LKSTGSFPPPV-----LEPLGASSETSKLOOLYEKDROGAVANASTASCAPTSA 616
Db ---VGGGRPPRPPSGPLSPRRPRAAGSPLPS-ATQIKOERAEUEYTESPAPRS 740
QY 617 PAPSSASGNQCIVLCRLVSCPALYLHGOHGGERPFCKVCGRATSTRONLPAHFVG 676
Db 741 PSEPPP-----KVVDVP-----SHASQSARENK 762
QY 677 HKTSPPARAONSRCPIOCKEFTTNAVTLLOOHVMILGGQIPNGSALISEGGAOENSSES 736
Db 763 H----LDRCNISCARSDLYF-----VPLEGSKLAKKRADILEVEVVRRA 801
QY 737 TASGPSFPQOQOPREPEEMSEDEEDEEDDYTDDEDISLAGRS--ESGEKAISV 793
Db 802 -----BORANEKEREREKEREKERKEBLESVKLAODGRAVESCISGRVNH 853
QY 794 RGDSEFEVSAGEEVATSVAARPVTUKEMDSNEKARONTLPRRPRRDNDIHPROMOGSTD 853
Db 854 RPPEE-PGSAVATVPRLYGBDPTALRTLSBXARNVH----SPGNRNHRYUVELGAUD 906
QY 854 VSGAMEEAKLEGISSMALLTOEGESTPTULVELINLRPMACKORPGESSGNKAEVCSQ 913
Db 907 -----FGLLGYNPVALYSSP----- 922
QY 914 SFTOTALENOHTPKDKGLFTCVFCROGFIDRALTKKNMLLANOV---PFAENG 968
Db 923 ----ZARENEREREND-----LRDRLKRGFEUVKPSLEBRJLVGVGPGLDPERRNG 969
QY 969 PONIATSLVPGCSSSTIPSPQLSFPP 994
Db 970 ----GLAIOPG-----PRELBEP 984
```

```

RESULT 20
B32891
finger protein 2, placental - human
C:Species: Homo sapiens (man)
C:Date: 20-Oct-1989 #sequence_revision 20-Oct-1989 #text_change 01-Dec-2000
C:Accession: B32891
R:Bellerfroid, E.J.; Lecocq, P.J.; Benhida, A.; Poncellet, D.A.; Belayew, A.; Martial,
DNA R, 377-387, 1869
A:title: The human genome contains hundreds of genes coding for finger proteins of th
A:Reference number: A32891; MUID:89377476; PMID:5505992
A:Accession: B32891
A:status: preliminary
A:molecule type: mRNA
A:Residues: 1-651 <BEL>
A:Cross-references: GB:M27878
C:Superfamily: zinc finger protein ZFP-36; LIM metal-binding repeat homology
C:Keywords: tandem repeat

Query Match          5.5%; Score 290; DB 2; Length 651;
Best Local Similarity 21.4%; Pred. No.9.8e-07;
Matches 135; Conservative 63; Mismatches 222; Indels 210; Gaps 20;

QY      364   EKPRGRHKRCFAKVFEGSDSALQIHLRSHTGEPKYCNVCNRRFTTGNLKVHFNHR-E 422
        ||| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB      173   EKP---YNCSQCCKAFKSQKSLGTSHQRTHTGKEPKYCEGGCKAFSRKSHLSHWRTHTGE 229

QY      423   KYPHVOMNPVPVEHLDYITSSGLPYGMSPPEKAEEEAAGTFGGVERKFLVASTIALS 482
        ||| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB      230   K-----PYGCN-----ECG-----RAFS 242

QY      483   ATESLTLLSTGTSTAARGLPTFENKFVLMAKEYPSKADENTPPSGSGSALAGVADSGSA 542
        :| : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB      243   EKSNI-----NHQRINTGKEPFECRECGKAPSRKSQ----- 274

QY      543   TRMOLSKIVTSIPSMWLTINHLKSTGSFPF-----PVLEPLGASPSSET 586
        ||| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB      275   -----LVTHIRHTHTGTRPFGSCDCKAKAPEKSELINRQTINTGKPYPC 318

QY      587   SKIQQLVEKIDROGAVAAVASTASGAPTTSAPADSSASGENOCVICLRVLCSPRLRLHY 646

```


406 RPTTRGLKLNHHNREKYPHYQAMPHPVPEHLDYVTTSSGLPYGMSVPPEKAEBAQIP 465
406 TSRTETRLKKHM---RES-----HTVEEQLRAGES-----EPAKESASSP 406
466 GGCVERRPLVASTALSTESTLLSTGTSTAVAPGLPTFNKFLMLKAVEPRSKADENIP 525
407 -----KNLSLKDGSAIT--SPINEINLSTTMASI----- 434
526 PSEGSATAGVADSGSATRMOLSKLVTSLPSWALLNLHLKSTGSPFPYVLEPLGASPE 585
435 LDSTNNNAV-----SSTSTEQPSAL-----SALTLDMSST-----PSSLSTLAHSSFG 477
586 TSKLQOLVEIDR-----QGAAVAVS---TASGAPITTSARPSSASAP----- 626
478 VSALDQIKATISENSEFMEPEGGINLALGCVSNAINKGDTPSPKQSNCECRSSSGKIRI 537
627 ---NQC-----VLCRLVISCPR-----LRLHYGGHGEREPFC 657
538 FKCKQCGHOSLKRDDMAHAKTHIPA EKQJNCOHCNFVTEKHHLEHYRHNIHIGSKPQC 597
658 KYCGRAFSTRGNLRAHFVGHKTPSPARAQNSCPICQK----- 694
598 KKCAVNCVKNMSMLNSHMSHTNHYQRCMD--CTYATKYCHSLKLHLKKYNHRRVPEGIEM 656
695 -----KFTNAVTLQ 703
657 SGGDSPPPTSDATITPESPLMKQEIKTETVEPYTSIAQPEPFENPMGNHGINFAMHLIN 716
704 OHVRMLUG-----QIPNGSALSSEGGAAOE-----NSSE--QSVASGSGSPQ 746
717 KHLDVGLMGLRNSVMSPLKCSACDFVASSADEKMRHSMHILNNSNPTSLASLYNSLNL 776
747 POSQOPSPSEEMSEEEDEDEEDVTDSDSLAGRSESGEKAISYRGDSSEEVSGAE-- 804
777 PEFSHVAPDNDMALESMDQVKRIDDDNITESHCEYEMDQSDSAVSPITGSSQISSGDEET 836
805 -----EEVATSVAAPTYTKEMDS--NEKAPQHTLPPPPPPDNLDPHPMEGTSDVS 855
837 KKCKSLSEIQISARANGNSPMNSDSEAMEKDCESADADAPHSPDPTTGVSPPLHSSSIYA 896
856 GAMEEPAKLEGISSEPMALPQEGECTSPPLVEELINPEAKKKDQEGESGRACEEVQGSF 915
897 PLPTTPQPNFFLOSILQASLIG-----PLL--ANRPSA-----YCDHCKIPF 938
916 PPTQALAEHQKTHPKDGPLFTCVFCRCQGLDRATLTKKMLLAHQ 960
939 DTQOVLDSHMRPHTPGND--FMCSDCQYQAFNELSPFLAHMYQARHQ 982

RESULT 23
31421
-terminal domain-binding protein RAI - rat
;Species: Rattus norvegicus (Norway rat)
;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 07-Dec-1999
;Accession: T31421
;YUYEV, A.; Patturajan, M.; Litingtung, Y.; Joshi, R.V.; Gentile, C.; Gebara, M.; Cord
;Title: The C-terminal domain of the largest subunit of RNA polymerase II interacts with
;Reference number: 221024; PMID:96293459; PMID:8692929
;Accession: T31421
;Status: Preliminary; translated from GB/EMBL/DBJ
;Molecule type: mRNA
;Residues: 1-1173 <YUR>
;Cross-references: EMBL:U49056; NID:g1438531; PID:g1438532; PIDN:AMC52657.1
;Experimental source: hippocampus

Query Match 5.4%; Score 286.5; DB 2; Length 1173;
Best Local Similarity 20.5%; Pred. No. 2.7e-06;
Matches 257; Conservative 109; Mismatches 418; Indels 467; Gaps 51;

79 PRPEGRSQVMDTEHNPDPSSGSGPPDPTMGPEKRGESSQGLVAATGTAAGCGGCL 138
59 PVPSTLPRRAWRGTGVSPQSHASRP-----ACSHHLLTIGT---GDGSPA 102

QY	139	ILASPKTGA--TPLEPESTAPPPPPPPPPGV-----	170
Db	103	PPAPSSGSSSPSPSSSSPPSPPPPPPPPPALPAPRPDIIDPHPTDEAYSPAPAE	162
QY	171	-----GSGHLNIPLLIEELVLOQRQHOMOTEOICROVLLGSL-----	211
Db	163	QKYDPEFATGSPSSSGGTPSPPEEEEE-----BEEEBEEELQSIRRISETLAGIYDONS	218
QY	212	-----GQVGAAPSSELPGGAASSTK-----	234
Db	219	LSQDFPEDDSEHRRPPPPQTLGAETPPPADSTRAGAPRRRVFVGPPEADACLEGKYSV	278
QY	235	-----PLPLEPSPIKPAQTGKTATSSSSSSSSSGAEPKQA--FPHLYHPLSQHPF	284
Db	279	EYVTTAGPALPL-PLPFP-----TDPEIEEGHIVGPEEPRAVSLFRAAR--RQPPA	330
QY	285	SVGVGSHKPT--PADSPALPGSTIDLIASPHLAPGTTGLLA-----AQCLGAR--	334
Db	331	SVATLSVAAPAAAPASAPRAPEGDFL--SLHADSDGEGALQVDLGEPPAPADAHWG	388
QY	335	GL-----EAAASPG-----LLKPNGSGELGYEVISSLEKGRHKC	372
Db	389	GDLRRKILTORRERYRORSASGPPPAKKARREKORSODPAPD--SPTWAKHRS	445
QY	373	RFCAKVFGSDSALOIHRSHTGERPKYKVCNCRFTRGNLKHFHRRREKYPVOMPH	432
Db	446	R--ERKJGSHSTARRSRSRRRS--RSMASAPR--RGS--HRSRSREKRRRRRSAS	496
QY	433	PVPEHLDVYTTSS-----GLPYMSVPEKAEABAG-----TPGGVE	470
Db	497	PPP-----AASSSSSRERHGRKREGKKKKKRSRAEKRSJGDELEKPAVPSSGD	551
QY	471	R-----KPLVSTALSTATESITLLSTGSTAVAPGLPTFNKFLMKAVEKSK---	519
Db	552	RSRRRGAVPPSIDLTDHDLFAIKRTTYVGRDKTEPPAPS-----PADVSGEVLX	606
QY	520	-----ADENTPPSESGAIGVADSGSATRQLSKLYT-----SLPSWALTLTNHL	564
Db	607	DSEGLADEHGAKGDKDORRRSGAASSSSSRKAKSRKALLDGBRGDRDSSKKPRPKD	666
QY	565	KSTGSPFPYVLEBLAGSPSETS-----KLQOLYEKIDROGAVAAVASTAGAPTTSPA	618
Db	667	SAPGSGALPKAPRPSGSSSSSSSSSKRVKLOSKAVLIREG--VST-----TPA	715
QY	619	PSGSASGPNQCVICLRLSCPRALRLHGGHGERFKKVCGRARSTRCNLRHFV---	675
Db	716	KDSSSSGLGS-----IYKFSRDRRESRPFLLPD	744
QY	676	-----GHKTSPPA-----RAONSCPIQOKKFTNAVTLQOHVRMHGGOIPNGSALSE	723
Db	745	ERSPAPAGVKAAPGSTRKKTAKAKAKAKAGKCTGKTPSKTRK-----KVRSGGSS7AS	800
QY	724	GG-GAAQEN-----SSDOSTASGPGSFPQPOSQOSP-----	754
Db	801	GGPGSLKKSADCSQAASAKGTEETSWSGERTTKAP--STPPKVAAPPALTPDSQTV	859
QY	755	-----DEEMSEEEDEDEEDDYDEDLAGSGSSGG	787
Db	860	DSSCKTPDVSYFLAEASBDPGVRYGAEEEBEEEEEBEEDQAPATTATSTAANASTA	919
QY	788	EKAISVYRGDSEEVYGAEEVATVSAAPTIVKEMDENKAPCHTLPPPPPPPDNDHPOPM	847
Db	920	PSAGSTVAGD--SGADGPRAARSQJLPL-----PPRPFWNL--PAGV	957
QY	848	EOGTSVSG-----AMEE-----EAKLEGJSSPMAALTOEGEGSTPLVEELINLPEAMK	896
Db	958	DOTTSVLTALLTALFKKEENALASRAKQOELLQDITNQIRHHRKPPSTLGTPAPVPTSFG	1017
QY	897	KDPGES-----GKKACEVCGSFTQJ-----	919
Db	1018	LPAPSSYLLPGSLPTGCGSSTPPTPGCLVPASDRGSSSSSEGRDGTDKYLKLHTQER	1077


```

119 SDPRDI--DQDNSTSPSI-XSPGSVENDSDSSSGLSQGARPRYHPPRLPPSPQPPST 175
101 -----GSSGPPDPPTWGPERRRGESSGQFLVA 126
176 PROPEASFEHPHSVTPRYGNAHMEPPTRMFOAPRGAPRPNPOLYPCGTGVLSPGPMGP 235
127 ANGTAAAGGGG-----LLASPKLGATPLPPEPSP-----APP----- 160
236 KGGGAASVGGPRGKONPPPTPTTSVSSGASGAPTKPTPTTVGGGNLPSPAPPANFP 295
161 -----PPPPGVS-----GHLNIPLLIEELRVLOOROIHQOMQTEOI 201
296 HTVPLNLPPLPALPLNNASAPPGIQAQPLPGLHLPSPY----- 333
202 CNOVLLIGSIGOTVG-----APASSELPGTGAASSTKPLLPLESPKPAQTC 249
334 -----AMGQMGGLPPEPEKGPETLAP-SPHSLP-----PASSAPAPPMFPYSSSSSS 381
250 KTTASSSSSSSSGAE--PRKQAFPHLYHPLGSHQHPFVSGVGRSHKPTAPSPALPST 307
382 SAAASSSSSSSSASAPPSQALPSPHSPPPPTSLV-----SNQPYTQPSLP--- 433
308 DOLIASPHLAPGTTGILLAAQCLGAARGLAAASDGLPKPKNGSGELGYGEVISLEKPG 367
434 SQAVMSQGP PPPPYGRLLA-----NSNAHGPFPPTGAQ-----STAHPPV 476
368 GHNKCFCAKVFSDSALQJHLRSHHTGERPYKCNCVGNRP--TTKGLKVHFRHREKYP 425
477 STHHHHHQQOQOQOQOQOQOHHNNGSGPP-----PGAFPHPLEGGSSHHAHYAMSPS 530
426 HVOHMPHY-PEHL-----DYVITSSGLPYGMSVPRPEKAEAEAGT----- 464
531 LGLIRYPRGPAHLP RHHSQVSYSQAGPNG--PRVSSSSNSSSTSSQSGTSPCHPSPSQ 587
465 -PGGVEKRRPLVASTALISATLSTLST-----GTSTAVAPGLPTFNK-----FVLMK 512
588 GQOQAPYPPPPYPTVTTSSATLS-TVIAVTASSPAGYKTASPPGPPYGRAPSPGAKT 646
513 AVERPSKABE-----NTPRGSEGS--AIAGVADSGSATRMOLSKVLTLSPWALLTNH 563
647 APPPGYKPGSPSPFRGTGPTGCTGPTGAPGTGPKGSP----- 686
564 LKSTGSFPPPY-----LEPLGASPESTSKLOOLVEKIDROGAVAVASGAPTSTA 616
687 ---VGRPLRPAGPSGLPRLPPRAAPASGFLS--ATQIKQPAEYETRPESVPPARS 741
617 PAPSSSASGPNQCVICIRVLSCPRALRLHYQONGGERPKCKVCGRASTRGNLRAHFG 676
742 PSPPP-----KVVDVP-----SHAQSAARNK 763
677 HKTSPARAARQNSPICOKKTTNATYTLQOHVKNHLGGQITPNGSALSEGGAAQNSSESQS 736
764 H---LDRFNSCARSDLYP--VPLE-----GPSWRSQPTWRRCCAGR----- 802
737 TASGGSFPQOQSPSEEMSESEDEEEDYTDSDSLAGRS--ESGGEKAISV 793
803 -----AARRRKCARARAEREKEREKERYERSVKLAQEGRAPVCEBGLGYPH 853
794 RGDSEEVGAEEVATVAAPTYVKEMDSNEKAPQHTLPPPPPNLDHROPMEQSTSD 853
854 RPPFE--PGSAVATVPYPLGPDPTPALRTLSEYARBYVM-----SPGNRNHPFYVPLGAVD 906
854 VSGAMEEELKLEISSPMALLTQEGEGSTPLVELINLPEAKKKDGESSGRKACEVCGQ 913
907 -----PGLIGYNVPAALYSSD----- 922
914 SFPQTALKEHOKTHPKDGLFTCVCFQOGLDRATLKKHMLLAHHQVP-----PFAPHG 968
923 ---AARERERARERD-----LRDLKPGFEVAKPSELEPLHGVPGGLDFFRHG 969
969 PONTATSLVPGCSSSIPSPGLSPFP 994

```

Db 970 -----GLAIQPG-----PPGLHPPF 984

RESULT 26

S35305

zinc finger protein ZNF91 - human

C:Species: Homo sapiens (man)

C:Date: 31-Dec-1993 #sequence_revision 02-Jun-1994 #text_change 05-Nov-1999

C:Accession: S35305

R:Belletroid, E.J.; Marine, J.C.; Ried, T.; Lecocq, P.J.; Riviere, M.; Amemiya, C.; P

EMBO J. 12, 1363-1374, 1993

A:Title: Clustered organization of homologous KRAB zinc-finger genes with enhanced ex

A:Reference number: S35305; MIMD:93223677; PMID:8467795

A:Accession: S35305

A:Molecule type: mRNA

A:Residues: 1-1191 <BEL>

A:Cross-references: EMBL:L11672; NID:9186773; PIDN:AAA59469.1; PID:9186774

A:Note: the authors translated the codon GCA for residue 750 as Thr and GCT for resid

C:Keywords: DNA binding; zinc finger

Query Match 5 3% Score 281; DB 2; Length 1191;

Best Local Similarity 21.98; Pred. No. 5e-06; Mismatches 245; Indels 176; Gaps 22;

Matches 137; Conservative 69; Mismatches 245; Indels 176; Gaps 22;

QY 364 EKPGRHKRCFAKVFSDSALQJHLRSHHTGERPYKCNCVGNRPPTRGMLKVFHHRREK 423

Db 627 EKP---YKCECGKAFSHSALAKHRIHTGEKPYKCECGKAFSSSTLANHKITHEE 683

QY 424 YPHVQMPHPVPEHNDYVTTSSGLPYGMSVPRKAE--EACPTGGVEKRPVASTAL 481

Db 684 KPYKCECDKTEKRLSTLTNKHIIHAG---EKLYKCECG-----KAF 723

QY 482 SATESITLSTGSTAVAPGLPTFNKFVLMKAVEPKSKADENTPPGSESGAIVGADSGS 541

Db 724 NRSSNLT-----HKFI--HTGEKPYCEE-----CGKAFNWS 754

QY 542 ATRMOLSKVLTSLS-----W-ALLTNHLK-STGSPFPYVLEPLGASPESTSK 588

Db 755 SSLTKHKRIHTREKPRCKCKGKAFIMSSFTLRHKRIHTGE--KPYKCECGKAFSSST 812

QY 589 LQOLVEKIDROGAVAVASGAPTSAAPASSASGNOCYTICLRVLSCPRALRLHYGQ 648

Db 813 LTK--HKTIHTG-----EKPYKCECGKAFKHSALAKHII 847

QY 649 HGERPFCCKVCGRAESTRGNLRAHFGHKTSPARAQNSPICOKKFTNNAVTLQOHVM 708

Db 848 HAGEKLYKCECGKAPNOSNLTKHIIHTKEPKSEED---CDKAFIMSSFTLTEHKRI 903

QY 709 HLGGQIPNGSALSEGGAAQNSSEGSTASGP-----GSFPOQOQSPSEEMSEE 761

Db 904 H-----TREKPYKCECGKAFSQPSHLTLTKHMTGEEK 936

QY 762 EEDDEEEDYTDSDSLAGRSSEGGKAISVKGDSSEVSGAEEVATVAAPTYKEMD 821

Db 937 PYKCECGKAFSSSTLTNKHIIHTGEKPYK-----EBCGKAFRSSSTLE-- 983

QY 822 SNEKAPQHTLPPPPPNLDHROPMEQSTSDVSGAMEEELKLEISSPMALLTQEGEGT 881

Db 984 -----HKTIH-----IGEKPYKCECGKAFSSSTLTTRITRNHTGE 1019

QY 882 STPLVEE---LNLPEAMKKDPSESSGRK--ACEVCGSFPQTALKEHOKTHPKDGLP 935

Db 1020 KPYKCECGKAFNRSSKLTTHKLIHTGEKPYKCECGKAFSSSTLNGHKRIHTREKP-Y 1078

QY 936 TCVCRCQGLDRATLKKHMLLAHHQVP 962

Db 1079 KCECGKAFSSSTLTNKHRIHTGEKP 1105

[illegible]

```

RESULT 32
06346
Inger protein (clone X1COE7.1) - African clawed frog (fragment)
;Species: Xenopus laevis (African clawed frog)
;Date: 28-Feb-1990 #sequence_rev1sion 28-Feb-1990 #text_change 01-Dec-2000
;Accession: S06546
;Mol. Biol. 208, 639-659, 1989
;Title: Second-Order repeats in Xenopus laevis finger proteins.
;Reference number: S05632; MUID:90040698; PMID:2509712
;Status: not compared with conceptual translation
;Molecule type: mRNA
;Residues: 1-615 <NIE>
;Superfamily: zinc finger protein ZFP-36; LIM metal-binding repeat homology
;Keywords: DNA binding; zinc finger

Query Match          5.0%; Score 266.5; DB 2; Length 615;
Best Local Similarity 21.6%; Pred. No. 1.2e-05;
Matches 135; Conservative 66; Mismatches 231; Indels 193; Gaps 25;

Y 364 EKPFGNKKCFKAVFSGSDSLQHLRSHGTGERPKYCNVCCGNRTTGNLKVHEHRREK 423
    |||
    |||
Y 140 EKP---FCSECKCFKFASSSELNIHQRTHTGKPFECSECKGCFTHNS---HFNHHQ-- 190
    |||
    |||
Y 424 YPHVQNPHPVPEHLVDVITSSGLPYGMSVPRPEKAEDEAGTPEGGVERRKPLVASTTALSA 483
    :||
    :||
Y 191 -----MHTG-----EKPCCSKGCKCF 208

Y 484 TESTLLSTGTSTAVADGLPTFNKFLMKAVEPKSRADNTPTPGSE--GSAIAGVADSSGA 542
    |||
    |||
Y 209 ASSSDI-----TFHR-----RTHRTKRTFCSECKGCF---SNHSHL 242

Y 543 TRMQL-----SKLVNLSLPSMALLTNHLKSTGSEFPPYVLEPLGASSESKL--Q 590
    :||
    :||
Y 243 ARHQMHHTGKEPKFCCSCGKCFSSSSGLTQHQRTHMKKVPFCSACGKCFNSRHLIRH 302

Y 591 OLVEKIDROGAVAVASTAGCAPTTSAPAPSSASGCPNOQYICRLVLSCPRALRLHYGQHG 650
    ||:
    ||:
Y 303 QMIH-----TGKPFSCPECKKCFSPNSLARKHQHTHT 335

Y 651 GERPKKVCAGRAFPSTRGNLRAHFVGKHTSPARAQNSCPICOKKFTNNAVTLQOHVRMHL 710
    |||
    |||
Y 336 GEKPFSCSCGKCFASSSDLETFH--HRTHTGEEK--PFSCSECKGCKYSSKSLVHHQRTHT 391

Y 711 GGQITNGSGALSGGGAAGQENSSSOSTASGPGSFQPGQSQQPFPEEMSEDEDEDEDEE 770
    |||
    |||
Y 392 -GEKPFSCSKDCDCEFASSSELNIHQRTHTGKFA-----SCSECKG 431
    |||
    |||
Y 771 DVTDDSDIAGRGSESGGEKAISVRGDSEE--VSG---AEEVATSVAAATPYKE---M 820
    |||
    |||

```

Db 432 CPTNSQLERHOMIHTEGKPISC-PECEECFVSSQLTAHQOQAHRMVKPFSCLECGKCF 490
 QY 821 DSNERAPQPTLLPPPPPPPPDLNDHPQBPHEOGSDVSGAMEEKEKLEGISSPMALITDEGES 860
 Db 491 SNRSNFARHOMIHTEGKPFSCSE---CRKGFNSQSLARHQ-----MTHTGE- 534
 QY 881 TSTPLVEELNLPBAMKKDPGESSGRKACEVGGSGFPQTALAEHQKTHPKDGLPTVCVC 940
 Db 535 -KPF-----SCSECGKRFFSNOSHARHOMIHTEGKPF-SCSECG 570
 QY 941 RQGLDRATLTKKMLLAHNVPPFA 965
 Db 571 AKGFNSQSLARHQ-MTHTEGKPEFA 594

RESULT 33
S26823

zinc finger protein ZNF43 - human
N.Alternate names: zinc finger protein kox27

C:Species: Homo sapiens (man)
C:Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 01-Dec-2000

C;Accession: S26823; I37967; S10416
R:Lovering, R.; Trowsdale, J.
Uncl. of Life Sci. 10 2001 2002 1001

A:Title: A gene encoding 22 highly related zinc fingers is expressed in ly
Nucleic Acids Res. 19, 2921-2928, 1991

A: Accession: S26823
A: Status: preliminary

A: Molecule type: mRNA
A: Residues: 1-803 <LOV>

A:Cross-references: EMBL:X59244; NID:g38031; PIDN:CAA41932.1; PID:g38032
R,Thiesen, H.J.

A>Title: Multiple genes encoding zinc finger domains are expressed in human

A:Accession: I37967
A:reference number: I37949; MUID:91145339; PMID:2268909
:Status: preliminary; translated from GR/EMBL/DDBT

```

A: Molecule type: mRNA
::ResIdues: 476-531 <THI>

```

A:Cross-references: EMBL:X52358; NID:g34160; PIDN:CAA36584.1; PID:g930090
C:Genetics:

A:Gene: GDB:ZNF43; HTF6
A:Cross-references: GDB:128653

A;Map position: 19p13.1-19p12
C;Superfamily: zinc finger protein ZFP-36; LIM metal-binding repeat homolo

Quarry Match 5 0%: Score 265. DP 2: Length 803.
c/keyscores: DNA binding; zinc finger

Match	Local Similarity	Score	DB	Indels	Gaps
Match 144: Conservative	20.8%	207	2	228	228
Match 145: Mismatch	1.9e-05	207	2	228	228
Match 146: Indel	1.9e-05	207	2	228	228

346 KPRKNGSGELGYGEVISTLEKPGGRHKCRFCAYFGSDSAIOIHLRSHGTGERPYKNCWGN 405

Db 287 KAFNOSSNLTEHKIHPGEKP---YKCECGGAFNWPSSLTKHKRHTTGKPKPTCECGCK 343

QY 406 RFTTRGNLKVHFHRH-----REKYPHYQMNPHVPPE----- 436

Db 344 AEFNFSNLTTHKRIHTAEKFKYKCEGGEAFSSSSNLTTHKKIHTEKKPYKCEEGKAFKK 403

QY 437 ---HDDVITSSGLPYGMSVPEPEAAEEAGTGGVERKPLVASTALSTATSLT---LL 490
..

Db 404 SSKLEHKLTHTG-----EKPYKOE-----CKAFNPSPILTKHNR 441

OY	491	STG	STAV	AP	GIP	-----	TEN	KE	-----	VLM	KAVE	PK	S	KAD	ENT	PP	SG	S	AI	AG	V	536
								:	:		:		:	:		:	:		:	:		

Db 442 HTGK-----PYKCEVCGKAFNDFSLTTHKRIMTAEKPYCEE-----CGK 483

```
QY 537 ADSSSATRMQLSKVLTVLSPWALLNHLKLSIGSFPPFVYVLEPLGASPSETSLQQLVERI 596
```

```
D8      484 AFSSNNLTKHKI-----HLEK-----PICEEGCAFRMSKUTE--IKI 524
```


[illegible]

```

OY 492 TGTSTAVAROLPFFNFKVLIMKAVERPESKADENTPRGSE-GSALAGADSGSATRMQLSKL 550
Db 73 TPR-----VARRHOKTHHTGKRPFCSCGKCFANSSD----- 104
OY 551 VTSLSFMAALLTNLTKSGSPFPYVLEPLGASPESTSKLOOLVEKIDROGAVAVASTAG 610
Db 105 -----VTIV-RRHHTGKRPYSCSGCKCFPTSSDL-----NVNRR----- 138
OY 611 APTTAPABSSASGNOCVLCLEVLSCPALRLHYGONGGERPRPKCVKAGRASTRGNL 670
Db 139 -----THTEKRPYSCSHCGKCFPTSSSELNVHRRHTHTGKRPYSCSGCKSFTPTSEF 189
OY 671 RAHFVGHKTPSARAQNSCPICQKFTNAVYLOQVNRHNLGGQIPRGSSALSSGGAAOE 750
Db 190 TSHWKTHMEKRP---SCVOGCKCFSKDTHLKHVYRHHTGK---PRSCFCGCKFTH 241
OY 731 NSS-----KOSTASGSGFPPOQSOQSPREEEMSEEEDEEEDVDDEDLAGR 781
Db 242 NSLSIKYHLKHKREADPCSKGNLNPET---SPVNHYPPTNKKEK-----AT 286
OY 782 GSESGGEKALSVNGDSEEVSGAEEEVATSVAAPTTVKEMDSNEKARQHTLPPRPDNL 841
Db 287 SWEEGNQSDYSINSLTEQIOGP-----YTPPTIMEYN-----HLI----- 321
OY 842 DHPQPEQGSVDVSGAANEKLEGISSPMALNLOEGESTRPLVE-----ELNLPEAMK 896
Db 322 -----MQDNKYDVNAC-----HSPL-----OETDTTKALNHRDIDRRORTOTOLK 362
OY 897 KDPGESSGRK-ACEVCGSFPTQTALDEHOKTHRPKDPGLFTVCFCROGFLRATLKKH 953
Db 363 YDHRTVTGDKPLFSCGCKCFSTYHVLARHQKTHNGEKP-FCSCEKCYARSSDLNVH 420

RESULT 39
S51037
zinc-finger protein - African clawed frog
C:Species: Xenopus laevis (African clawed frog)
C:Date: 01-Aug-1995 #sequence_revision 01-Sep-1995 #text_change 01-Dec-2000
C:Accession: S51037; S06565
R:Schaefer, U.; Rausch, O.; Boumeester, T.; Pieler, T.
Eur. J. Biochem. 226, 567-576, 1994
A:Title: Sequence-specific recognition of a repetitive DNA element by a C(2)H(2) zinc
A:Reference number: S51037; MUID:95094815; PMID:8001572
A:Accession: S51037
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-675 <SCH>
R:Nietfeld, W.; El-Barddi, T.; Mentzel, H.; Pieler, T.; Koester, M.; Poeting, A.; Kno
J. Mol. Biol. 208, 639-659, 1989
A:Title: Second-order repeats in Xenopus laevis finger proteins.
A:Reference number: S05632; MUID:90040698; PMID:2505712
A:Accession: S0565
A:Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 85-448, 'K', 'A', 450-613 <NIE>
A:Title: Experimental source: clone X1CGF20-1
C:Superfamily: zinc finger protein zfp-36; LIM metal-binding repeat homology
C:Keywords: DNA binding; zinc finger

Query Match 4.9%; Score 261.5; DB 2; Length 675;
Best Local Similarity 20.1%; Pred. No. 2.3e-05;
Matches 128; Conservative 75; Mismatches 255; Indels 179; Gaps 21;

OY 364 EKRGGNHKCRFCAKVVGSDBALQIHLNLSHTGERPRYKAVCGARFTTRGLYKHFNHNRK 423
Db 143 EKP---YSCQCGKSFANHCYLDSSHQRTHTGDKPRSCYECGCKFFSQRGNLNHLTKTKLD 199
OY 424 YPH-----YQNPNRVPENHIDYVITSGLLYGMSVPRPEKAEDEAG 463
Db 200 QPHLACGCKTFPKSTGLLENQKHSELRPLSECGKTTSDAHNLKHKOSTPE----- 252
OY 464 PRGGVERKPLVASTA-LSATESLLTLLSTGTSVAVADGLTFENKFLMKAVEPKSKADE 522

```



```

b 253 -----EQRPRCTEGCEIFSENEHELLTHOSTHTEDEOKPRPPTCWMGIF-SNEHEHRTQ 305
y 523 NTPPGSEGAIAVADSGSATSARMOLSKIVTSLPSWMLTLTNLTKST--GSPFPVYLEDL 579
b 306 ST--HNEGQSLPSTESG-GTFSENEHELLTHOSTHNEGKHLRPTCEGCTFTNEG--ELL 360
y 580 GASPSSTSKIQOLV-----EKIDROGAVAVASTASGAPTT----SAPADSSA 623
b 361 AHOSTHTEDEOKPRPCTEGCEIFSDHEHELLTHOSTHSPSPSEFGVQTTEDNHQSPSKDHG 420
y 624 SGPNOCVICRLVLSCPRALRYLHYGONGGERPRFKCYKXGAFSTGYNLRAHFNGHTSPAA 683
b 421 EKPRPSCSECKSPFYKSVLKDHLYVHTGEKPYHCIECGKSTYHOSSLSKH--QRHTGCV 477
y 684 RAONSCPICOKKFTNAVTLQOHVRMLAGQIPINGSALSSEGGAOENSSEOSTASPGS 743
b 478 KA-FSCNLCD-KLIIISKLRLHYRHSG----- 503
y 744 PPOPOQOPSPREEEMSEBEDEDEEDVTDEDSLAGRGSBGEKALISVKGDSERVSQA 803
b 504 -----EKRPCTEC-----DKPTKKEQLESHYKVTHTGEK----- 533
y 804 EEEVATSVAPRTYTKEMDSNEKARQHTLPRPPRPDLDBPQMEQGTSDVS--GAMEEE 861
b 534 -----PYPOQCGKSPFHKSVYLAH 553
y 862 ANLEGISSPMAALTOEGEGSTPLVEEELNLPEAMKKDPGSESGRKACVSGOSFPOTAL 921
b 554 LRTHTDOKPRPSC-TECGTFTTKRPVYESHLTTHTKGRP-----FSCIECGKEFMAKRN 606
y 922 EENOKTHPKDGRPLFTCVFCROGEFLDRATLKKHMLAH 958
b 607 EAHYKMHTEGEP-FTTECGKFTFTWKSNNLSRSHYTTVAH 642

```

```

RESULT 40
NMHUE2
HIV-EP2 enhancer-binding protein - human
Accession: S26661, A38829, A38253
Date: 31-Mar-1993 #sequence_revision 07-Jul-1995 #text_change 21-Jul-2000
Species: Homo sapiens (man)
Title: Structure and expression of major histocompatibility complex-binding protein 2
Accession: S26661
Status: nucleic acid sequence not shown
Molecule type: mRNA
Cross-references: EMBL:X65644; NID:q38259; PIDN:CAA6596.1; PID:q38260
Biochem. Biophys. Res. Commun. 266, 8590-8544, 1991
Title: HIV-EP2, a new member of the gene family encoding the human immunodeficiency virus
Reference number: A39829; MUID:91217105; PMID:2022670
Status: nucleic acid sequence not shown
Molecule type: mRNA
Residues: 668-2144, R, 2146-2500 <NOM>
Cross-references: GB:M60119; NID:g2661140; PIDN:AAB88218.1; PID:g162120
Rustgi, A.K.; Van't Veer, L.J.; Bernards, R.
Proc. Natl. Acad. Sci. U.S.A. 87, 8707-8710, 1990
Title: Two genes encode factors with NF-kappaB- and H2TF1-like DNA-binding properties
Reference number: A38253; MUID:91062349; PMID:2247438
Accession: A38253
Molecule type: mRNA
Residues: 1851-1990 <RUS>
Cross-references: GB:M61744; GB:M33920; NID:g187404; PIDN:AAA36202.1; PID:g187405
Genetics:
Gene: GDB:HIVER2
Cross-references: GDB:129086; OMIM:143054
Map position: 6q23-6q24

```

C:Superfamily/HTV-Ep2 enhancer-binding protein
C:Keywords: DNA binding; duplication; metal binding; transcription regulation; zinc f
F:242-298/Region: DNA binding #status predicted
F:245-265/Region: zinc finger CCHH motif
F:273-295/Region: zinc finger CCHH motif
F:991-997/Region: nuclear location signal
F:1004-1036/Region: serine-rich
F:1852-1908/Region: DNA binding #status predicted
F:1855-1875/Region: zinc finger CCHH motif
F:1883-1905/Region: zinc finger CCHH motif
F:1953-1977/Region: acidic

Query Match	4.9%;	Score 261.5;	DB 1;	Length 2500;
Best Local Similarity	21.4%;	Pred. No. 9.5e-05;		
Matches 229; Conservative	97;	Mismatches 352;	Indels 391;	Gaps 52;

QY	243	IKPAOTGKTTASSSSSSSSSGAEP-----PKOAFPHLPHLGLSGHPSV-----	CG	288
Db	104	IEPBOIG-NTASAQLEFGSGKLASPEVYQOVAEKQYPPHPRSPYSCHSHSFPQHSILPOG	162	
QY	289	VGRHKRPAPASPALPGSTDQLIASPHLAFPGTTLAQAACGARGLEAASPGILK-P	347	
Db	163	VMHSTKP-----HQSLGPPWLPBGP-----LPVASEDLFPFP	196	
QY	348	KNG-SGELGY-GEVITSLP-----KPGRHKCFCAKVFGS	381	
Db	197	IHGHSQ-GYPRKKITSSLINPAYQOYSOKSIOAEAEHAKKHKPKPKGKICOPYCSRAAC	254	
QY	382	DSAQIHRSHTGEPYKCNVCGNRFTRGMLKVPHRHREKPHV-----	427	
Db	255	PSYVAKHHRSHTGEPYPCIPCGSFEKTSNL-----YKRRKSAHAIAKAGLVPTTESAVS	310	
QY	428	-----QNNPHVPPEHLDVYITSSGLPYG	450	
Db	311	KLDLEAGRIDWEAETHSDQSTDTDESSLFAEASDKNSPG-PILPD--IASGQYHG	367	
QY	451	MSVPEKAEERAG-----TPGGVERKPLVASTALSTATESLTLLSTGSTAIVAPG	501	
Db	368	-----SLEESIGCGPMKVPLILIPKSGI--PLPRESSOYIGPDMLPNPSLNTX--ADD	415	
QY	502	LPENKFLPMKAVEPKSKADEN-----PQSGESAIAGVAADSGSATPMOLSKLYLSW	557	
Db	416	SHYVKQKALRLSEKKGQOSEPSLNLISPHSKGSDSGYFSNSESABQOISPPMINKASY	475	
QY	558	ALL-----TNLKSTGSF-----PFYVLEPL-----	579	
Db	476	ELIIFGKCYRLSPRNALSTYTTQGERAMGRKGIMEPLRHVNRLDVKMFEDPVSQLPS	535	
QY	580	--GASPSETSKEI-----QOLVEKIDROGAVAAVASTASG-----APTTSAP--	617	
Db	536	KGVDPQSTSMKSTKFENSESROPQIIPSSIRNEKILYPAFQGSNVLLEAPVDYSSPL	595	
QY	618	-----APSSASGPNQCVICIRLVSCPRALRLHYGH-----GGERPFKCKVG--	661	
Db	596	RNSVPTISATN-----LTIIPSLR--GSHSEPERMTGSDVFTYGTGVIIPOR	642	
QY	662	-----RAFS-----TRGNLRAHFVGH-----KTSPARAONSCPIQ--KKF	696	
Db	643	MLRQAAAEPLPSVQEBHVEVHHGMLKGISSSLKEKTLSPDDRVGVDYDVCRRPYKMW	702	
QY	697	TNAVTLQOHR-----MHLGO-----IPNGSALSEGGAOENS--SEOS	736	
Db	703	EDSETPKQNYRDISCLSLKHGGEYFMDPVPLQGVPSMFGTTCENKRRREKSVGDEED	762	
QY	737	T-----ASGGSF-----POPOSOQ-----PSPEEMSEEBEED	765	
Db	763	TPMICSISTVSPVGIWASDYDPKIQMGEGVRSGFAMAGHENLHGHTEREDPCRPQLOPG	822	
QY	766	-----EEEEEVDYDDESLAGSGSESAGEKAISV-----RDSSEEVSGAEEEVA-TSY	811	
Db	823	SPSLVSESPSAIDSKMMDLGGRRKPPGCVNIVSYIGHTNSLSRNSPFRSEBAELVACTOD	882	
QY	812	AAPTTVKEMDSN-EKAPQHTLPPPPPPDNL--HPQMEQGTSDVSGAMEEAKLEGI	867	

Db 883 KAPSPSETCDSEISEAP--VSPEWAPPDGAESGGKPSPSQVQOOSYHTQPRLVHQNI 940
QY 868 SSPMALIQEGEGTSTPL-----VELNLP---EAMKKDPGE----- 901
Db 941 QVPEIRVTEEPDPKEKEKAQSKPEKPEVEEFOWPQSETLSQLPAEKLPPKKRRLIAD 1000
QY 902 ---SSGRKACEVCGOSF---PTOTALEEHOKTHPKDGPLFTCVFCROGFLDRATLKKHML 955
Db 1001 MEHSSGESSEFESTGTGLSRSPSQESNLSSHSS-----FSMSFEREETSKLSALPKODE 1053
QY 956 LAHQVPPFAPHPGPNITATLSLVPG-----CSS-SIPSPGLSPFP 994
Db 1054 FGKHSEFLTVPAGSYSL-----VPGHHHQKEMRCCSSBQMPCPHPAEVP 1098

Search completed: January 13, 2003, 15:21:29
Job time : 54.4664 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 13, 2003, 15:17:40 : Search time 11.4828 Seconds
(without alignments)
3619.260 Million cell updates/sec

Title: US-09-988-117-3
Perfect score: 5301
Sequence: 1 MAOETGSSRLGPGCEPAE.....SSIPSPGLSPFPKDDPTMP 1002

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5
Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	5182	97.8	1004	1	SAI2_MOUSE
2	4618.5	87.1	1007	1	SAI2_HUMAN
3	1338	25.2	1300	1	SAI3_HUMAN
4	1301	24.2	1324	1	SAI1_HUMAN
5	1282	24.0	1323	1	SAI3_MOUSE
6	1273.5	24.0	1322	1	SALI_MOUSE
7	759.5	14.3	1355	1	SALM_DROME
8	711.2	13.4	1402	1	SALM_DROME
9	371.5	7.0	1845	1	SALM_DROME
10	323.5	6.1	773	1	ZN41_HUMAN
11	311	5.9	734	1	ZN42_HUMAN
12	306.5	5.8	654	1	ZN43_HUMAN
13	303	5.7	722	1	ZN44_HUMAN
14	301	5.7	1829	1	ZN45_HUMAN
15	290	5.5	738	1	ZN46_HUMAN
16	290	5.5	1185	1	DRPL_HUMAN
17	290	5.5	1350	1	XFTN_XENLA
18	287.5	5.4	982	1	XFTN_XENLA
19	283.5	5.3	2688	1	ZEPI_MOUSE
20	282.5	5.3	818	1	ZEPI_MOUSE
21	281	5.3	1191	1	KR18_HUMAN
22	278.5	5.3	512	1	ZN91_HUMAN
23	274	5.2	706	1	Z151_CHICK
24	271.5	5.1	898	1	Z071_XENLA
25	271	5.1	1060	1	ZFHL_DROME
26	270.5	5.1	780	1	Z084_XENLA
27	270.5	5.1	839	1	ZF95_XENLA
28	270	5.1	882	1	ZF95_HUMAN
29	269	5.1	1042	1	EVIL_MOUSE
30	268.5	5.1	717	1	PRDD_HUMAN
31	268	5.1	697	1	Y441_HUMAN
32	267	5.0	578	1	Z192_HUMAN
33	267	5.0	683	1	Z263_HUMAN

34	266	5.0	845	1	KR0H_DROME	P08155 drosophila
35	265	5.0	803	1	ZN43_HUMAN	P17038 homo sapien
36	263.5	5.0	686	1	ZN07_HUMAN	P17097 homo sapien
37	263	5.0	803	1	Z151_HUMAN	O13105 homo sapien
38	263	5.0	1167	1	Z208_HUMAN	O43345 homo sapien
39	262	4.9	604	1	GLAS_DROME	P13360 drosophila
40	261.5	4.9	645	1	ZF93_MOUSE	O61116 mus musculu
41	261.5	4.9	759	1	Z287_MOUSE	O9e9b9 mus musculu
42	261	4.9	428	1	ZF26_MOUSE	P10076 mus musculu
43	260.5	4.9	1342	1	Z335_HUMAN	O9h422 homo sapien
44	260	4.9	594	1	ZF37_MOUSE	P17141 mus musculu
45	260	4.9	751	1	Z184_HUMAN	O99676 homo sapien
46	259.5	4.9	803	1	Z226_HUMAN	O9hnt6 homo sapien
47	259.5	4.9	1183	1	DRPL_RAT	P54258 rattus norv
48	259	4.9	553	1	Z324_HUMAN	O75467 homo sapien
49	258.5	4.9	751	1	Z337_HUMAN	O9y3m9 homo sapien
50	258.5	4.9	794	1	Z151_MOUSE	O60821 mus musculu
51	257.5	4.9	675	1	ZG20_XENLA	P18714 xenopus lae
52	256	4.8	453	1	Z06_XENLA	P18749 xenopus lae
53	256	4.8	2161	1	SHK1_HUMAN	O9y566 homo sapien
54	255.5	4.8	720	1	PRDH_HUMAN	O9h9d4 homo sapien
55	254.5	4.8	595	1	ZN85_HUMAN	O03923 homo sapien
56	254.5	4.8	754	1	Z287_HUMAN	O9hbt7 homo sapien
57	254	4.8	439	1	Z028_XENLA	P18747 xenopus lae
58	252.5	4.8	670	1	ZN16_HUMAN	P17020 homo sapien
59	252	4.8	2167	1	SHK1_RAT	O9ww48 rattus norv
60	251.5	4.7	1163	1	Y222_HUMAN	O92618 homo sapien
61	251.5	4.7	1258	1	GLI2_HUMAN	P10070 homo sapien
62	250.5	4.7	477	1	MAZ_HUMAN	P36270 homo sapien
63	249	4.7	616	1	ZN93_HUMAN	P35789 homo sapien
64	249	4.7	739	1	YD04_HUMAN	O9p2j8 homo sapien
65	248	4.7	476	1	Y106_HUMAN	O961j9 homo sapien
66	248	4.7	595	1	Z317_HUMAN	O9dp66 homo sapien
67	247.5	4.7	477	1	MAZ_MOUSE	P56671 mus musculu
68	247.5	4.7	783	1	ZF25_HUMAN	O9u1j5 homo sapien
69	247	4.7	892	1	HICI_MOUSE	O941y5 mus musculu
70	247	4.7	947	1	Z268_HUMAN	O14587 homo sapien
71	246.5	4.7	913	1	Z228_HUMAN	O9u1j3 homo sapien
72	246.5	4.7	1051	1	EVIL_HUMAN	O03112 homo sapien
73	246	4.6	589	1	ZN44_HUMAN	P15621 homo sapien
74	245.5	4.6	469	1	Z135_HUMAN	P52738 homo sapien
75	245.5	4.6	755	1	RREI_HUMAN	O92766 homo sapien
76	244	4.6	510	1	ZF64_HUMAN	P15622 homo sapien
77	244	4.6	614	1	ZF29_MOUSE	O07230 mus musculu
78	244	4.6	821	1	ZN41_HUMAN	P51814 homo sapien
79	244	4.6	1276	1	PRDC_HUMAN	O9haz2 homo sapien
80	244	4.6	2717	1	ZEPI_HUMAN	P15822 homo sapien
81	243	4.6	582	1	ZF36_HUMAN	P16415 homo sapien
82	242	4.6	488	1	Z345_HUMAN	O14585 homo sapien
83	241.5	4.6	354	1	KLF2_MOUSE	O60843 mus musculu
84	241	4.5	698	1	Z234_HUMAN	O14588 homo sapien
85	240	4.5	351	1	KLF2_RAT	O9e158 rattus norv
86	240	4.5	457	1	Z140_HUMAN	P52738 homo sapien
87	240	4.5	627	1	Z264_HUMAN	O43296 homo sapien
88	239.5	4.5	688	1	HKR3_HUMAN	P10074 homo sapien
89	239	4.5	509	1	ZKRI_CHICK	P30373 gallus gall
90	239	4.5	636	1	ZF90_MOUSE	O61967 mus musculu
91	239	4.5	819	1	ZF95_MOUSE	O924d8 mus musculu
92	239	4.5	1687	1	Z142_HUMAN	P52746 homo sapien
93	238.5	4.5	803	1	ZXDB_HUMAN	P98169 homo sapien
94	238.5	4.5	3866	1	HRX_MOUSE	P55200 mus musculu
95	238	4.5	601	1	ZF37_RAT	O88553 rattus norv
96	237.5	4.5	733	1	HICI_HUMAN	O14526 homo sapien
97	237	4.5	1029	1	Z197_HUMAN	O14709 homo sapien
98	237	4.5	1403	1	CIC_DROME	O9u1h0 drosophila
99	237	4.5	3703	1	ABFI_HUMAN	O15911 homo sapien
100	236.5	4.5	614	1	ZF28_MOUSE	P10078 mus musculu

ALIGNMENTS

RESULT 1

AL2_MOUSE
D SAL2_MOUSE STANDARD; PRT; 1004 AA.
C 090X96;
T 15-JUN-2002 (Rel. 41, Created)
T 15-JUN-2002 (Rel. 41, Last sequence update)
T 15-JUN-2002 (Rel. 41, Last annotation update)
T Sal1-like protein 2 (Spalt-like protein 2) (MSal-2).
S SAL12 OR SAL2.
S Mus musculus (Mouse).
S Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
C Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
C NCBI_TaxID=10090;
P [1]
P SEQUENCE FROM N.A.
P TISSUE-Brain:
P MEDLINE=20069074; PubMed=10602995;
P Kohlhaase J., Altman M., Archangelo L., Dixkens C., Engel W.;
T "Genomic cloning, chromosomal mapping, and expression analysis of
L Msal-2.";
L Mamm. Genome 11:64-68(2000).
C -1- FUNCTION: Probable transcription factor.
C -1- SUBCELLULAR LOCATION: Nuclear (Probable).
C -1- TISSUE SPECIFICITY: Expressed throughout embryonic development. In
C adult predominantly in brain.
C -1- SIMILARITY: BELONGS TO THE SAL FAMILY OF C2H2-TYPE ZINC-FINGER
C PROTEINS.
C -----
C This SWISS-PROT entry is copyright. It is produced through a collaboration
C between the Swiss Institute of Bioinformatics and the EMBL outstation -
C the European Bioinformatics Institute. There are no restrictions on its
C use by non-profit institutions as long as its content is in no way
C modified and this statement is not removed. Usage by and for commercial
C entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
C or send an email to license@isb-sib.ch).
C -----
C EMBL; AJ007396; CAB65274.1; -.
C MGD; MG1:1354373; Sal12.
C InterPro; IPR000822; Znf_C2H2.
C Pfam; PF00096; zf-C2H2; 7.
C SMART; SM00355; ZNF_C2H2; 7.
C PROSITE; PS00028; ZINC_FINGER_C2H2_1; 7.
C PROSITE; PS50157; ZINC_FINGER_C2H2_2; 7.
C Transcription regulation; DNA-binding; Nuclear protein; Zinc-finger;
W Metal-binding; Repeat.
T ZN_FING 372 394 C2H2-TYPE.
T ZN_FING 400 422 C2H2-TYPE.
T ZN_FING 629 651 C2H2-TYPE.
T ZN_FING 657 679 C2H2-TYPE.
T ZN_FING 689 711 C2H2-TYPE.
T ZN_FING 908 930 C2H2-TYPE.
T ZN_FING 937 961 C2H2-TYPE.
T ZN_FING 135 139 POLY-GLX.
T DOMAIN 135 170 POLY-PRO.
T DOMAIN 158 170 POLY-SER.
T DOMAIN 256 264 POLY-GLU.
T DOMAIN 762 772 POLY-GLU.
T DOMAIN 834 840 POLY-PRO.
Q SEQUENCE 1004 AA; 104975 MW; E46DD2C9C26A4C92 CRC64;

QY 196 OMTEQICROVLLLSGQTVGAPASPELPTGAASSTRPLLPSPIRPAOTGKTASS 255
C |||||
Db 198 QMTEQICROVLLLSGQTVGAPASPELPTGAASSTRPLLPSPIRPAOTGKTASS 257
QY 256 SSSSSSGAEPPEKAFPHLVLPLGSGHPFVGVGVRSHKPTPAPSPALPGSTDLIASPH 315
C |||||
Db 258 SSSSSSGAEPPEKAFPHLVLPLGSGHPFVGVGVRSHKPTPAPSPALPGSTDLIASPH 317
QY 316 LAFPGTTGLLAQCLGAARLEAASPGLLKPKNGSGELGYEEVYSSLRKPGGRHRCFC 375
C |||||
Db 318 LAFPGTTGLLAQCLGAARLEAASPGLLKPKNGSGELGYEEVYSSLRKPGGRHRCFC 377
QY 376 AKVFGSDSALQIHLRSHTGERPKCNVCGNRFETTGKLVHHRHREKYPHYVMNPAPVP 435
C |||||
Db 378 AKVFGSDSALQIHLRSHTGERPKCNVCGNRFETTGKLVHHRHREKYPHYVMNPAPVP 437
QY 436 EHLDEVITSSGLPYGMSVPEPKAEELAGTPGGGVERRKPLVASTALSTATESLTLSTGTS 495
C |||||
Db 438 EHLDEVITSSGLPYGMSVPEPKAEELAGTPGGGVERRKPLVASTALSTATESLTLSTGTS 497
QY 496 TAVAGLPTFNKFLVLMKAVPKSKADENTPPGSESSALAGVADSGATRMQLSVLSLP 555
C |||||
Db 498 TAVAGLPTFNKFLVLMKAVPKSKADENTPPGSESSALAGVADSGATRMQLSVLSLP 557
QY 556 SMALLTNHLKSTGSPFPYVLEPLGASPSSETSKLOQLVEKIDROGAVAAVASTASGAPTS 615
C |||||
Db 558 SMALLTNHLKSTGSPFPYVLEPLGASPSSETSKLOQLVEKIDROGAVAAVASTASGAPTS 617
QY 616 APAPSSASGPNQCVICLRVLCSPRALRLHYGQHGGERPFKCKVCGRAFSTRGNLRAHV 675
C |||||
Db 618 APAPSSASGPNQCVICLRVLCSPRALRLHYGQHGGERPFKCKVCGRAFSTRGNLRAHV 677
QY 676 GHKTSPPAARAQNSCPICQKFTNVAVTLOQHVMHGGQIPNGSGALSSEGGAAQENSSEQ 735
C |||||
Db 678 GHKTSPPAARAQNSCPICQKFTNVAVTLOQHVMHGGQIPNGSGALSSEGGAAQENSSEQ 737
QY 736 STASGPGSPFOQSOQSPSPPEEKSEEEDEEEDVDDEDLAGSSESGEKALISVRG 795
C |||||
Db 738 STASGPGSPFOQSOQSPSPPEEKSEEEDEEEDVDDEDLAGSSESGEKALISVRG 797
QY 796 DSEEVSGAEEVAATVAAPTVCVEMDSNEKAPQHTLPPEPPPPNDIDHPQPMQGTSDVS 855
C |||||
Db 798 DSEEVSGAEEVAATVAAPTVCVEMDSNEKAPQHTLPPEPPPPNDIDHPQPMQGTSDVS 857
QY 856 GAMEEAKLEGITSSPMAALTQDEEGTSTPLVEELNLPEAMKRDPCSSSGKACEVCGQSF 915
C |||||
Db 858 GAMEEAKLEGITSSPMAALTQDEEGTSTPLVEELNLPEAMKRDPCSSSGKACEVCGQSF 917
QY 916 PTQALAEHQKTHPKDGPLFTVCVFCRCGQFLDRATLKKHMLAHHQVPPRAPHGPQNTATL 975
C |||||
Db 918 PTQALAEHQKTHPKDGPLFTVCVFCRCGQFLDRATLKKHMLAHHQVPPRAPHGPQNTATL 977
QY 976 SLVPGCSSSIPSPGLSPFPRKDDPTMP 1002
C |||||
Db 978 SLVPGCSSSIPSPGLSPFPRKDDPTMP 1004

Query Match: 97.8%; Score 5182; DB 1; Length 1004;
Best Local Similarity 99.3%; Pred. No. 1.4e-216;
Matches 980; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

RESULT 2
SAL2_HUMAN
ID SAL2_HUMAN STANDARD; PRT; 1007 AA.
AC 09Y467; 09Y4G1;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DE Sal1-like protein 2 (Zinc finger protein Sal12) (Hsal12).
GN SAL12 OR SAL2 OR KIA0360.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OC NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 1).

RX MEDLINE=97131507; PubMed=89757505;
RA Kohlhaase J., Schuch R., Dowe G., Kuehnlein R.P., Jaeckle H.,
RA Schroeder B., Schulz-Zweiffer W., Kreisrman H.A., Koehler A.,
RT Mueller U., Raab-Vetter M., Burhardt E., Engel W., Stick R.,
RT "Isolation, characterization, and organ-specific expression of two
novel human zinc finger genes related to the Drosophila gene spalt.";
RL Genomics 38:291-298(1996).
RN [2]
RC TISSUE=Brain;
RP SEQUENCE OF 11-1007 FROM N.A. (ISOFORM 2).
RZ MEDLINE=97349984; PubMed=9205841;
RA Nagase T., Ishikawa K.-I., Nakajima D., Ohira M., Seki N.,
RA Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
RT Prediction of the coding sequences of unidentified human genes. VII.
The complete sequences of 100 new cDNA clones from brain which can
code for large proteins in vitro.";
RL DNA Res. 4:141-150(1997).
RN [3]
RP SEQUENCE OF 541-1005 FROM N.A. (ISOFORM 2).
RZ Morgan J.W., Ford D., Ma Y., Matzel A.L.;
RA "Homo sapiens mRNA for zinc finger protein, SALL2 exon 2.";
RT Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
CC CC
CC - FUNCTION: Probable transcription factor.
CC - SUBCELLULAR LOCATION: Nucleus (Probable).
CC - ALTERNATIVE PRODUCTS: 2 isoforms; 1 (shown here) and 2; may be
produced by alternative splicing.
CC - TISSUE SPECIFICITY: Highest levels in adult brain (in different
areas). Lower levels in heart; very low levels in kidney and
pancreas.
CC - DEVELOPMENTAL STAGE: In fetal brain exclusively in pontine nuclei.
CC - SIMILARITY: BELONGS TO THE SALL FAMILY OF C2H2-TYPE ZINC-FINGER
PROTEINS.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)
or send an email to license@isb-sib.ch).

EMBL; X98834; CAA67331.1; ALT_INIT.
DR EMBL; AB002358; BAA21638.1; -.
DR EMBL; AF465630; BAL74188.1; -.
DR Genew; HGNC:10526; SALL2.
DR MIT; 602219; -.
DR InterPro: IPR000822; ZnF_C2H2.
DR Pfam; PF00096; Zf-C2H2; 7.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 7.
DR PROSITE; PS0157; ZINC_FINGER_C2H2_2; 7.
KM Transcription regulation: DNA-binding; Nuclear protein; Zinc-finger;
Metal-binding; Repeat; Alternative splicing.

FT FT
FT ZN_FING 373 395 C2H2-TYPE.
FT ZN_FING 401 423 C2H2-TYPE.
FT ZN_FING 631 653 C2H2-TYPE.
FT ZN_FING 659 681 C2H2-TYPE.
FT ZN_FING 691 713 C2H2-TYPE.
FT ZN_FING 911 933 C2H2-TYPE.
FT ZN_FING 940 964 C2H2-TYPE.
FT DOMAIN 135 139 POLY-GLY.
FT DOMAIN 158 170 POLY-PRO.
FT DOMAIN 256 265 POLY-SER.
FT DOMAIN 764 775 POLY-GLU.
FT DOMAIN 816 821 POLY-ALA.
FT DOMAIN 838 843 POLY-PRO.
FT VARSPIC 575 581 LPLCARA -> PPVYLEP (IN ISOFORM 2).
FT CONFLICT 11 24 QLSIDGSPASSEN -> SRLGVPCGPAPELG (IN
REF. 2).
FT CONFLICT 547 547 L -> R (IN REF. 1).
FT CONFLICT 554 554 V -> M (IN REF. 1).
SO SEQUENCE 1007 AA; 105135 MW; 9209963225A5459E CRC64;

Query Match 87.1%; Score 4618.5; DB 1; Length 1007;

	Best Local Similarity	89.1%	Pred. No. 2.6e-192;	
	Matches	882;	Conservative	23;
			Mismatches	82;
			Indels	3;
			Gaps	3;
QY	16	GPABERGSGASBEHHNPVCAKCCAKQSPPTPEFLAHNONSCTDPRVWVITIGGOENPNSSA	75	
Db	18	GPSASGNGASDEHDHPVCAKCCAKQFTDPTPEFLAHNONSCTDPRVWVITIGGOENPNSSA	77	
QY	76	SSAPREGHSHRSQVMDTETHSNRPDSSGSGPRDPITWGPERRGESSGQFLVAATGAAGG	135	
Db	78	SSEPREGHNPNQVMDTETHSNRPDSSGSGVPLDPTWGPERRGESSGHFLVAATGAAGG	137	
QY	136	GGLIASPXLGATPLPEESTPAPRPPPPPPPPGVGSHLNPILIELRLVLOQROIHQ	195	
Db	138	GGLIASPXLGATPLPEESTPAPRPPPPPPPPGVGSHLNPILIELRLVLOQROIHQ	197	
QY	196	QMTBOTCRVLLIGSGOTVGAPASPELPESTGAASSRKPLPLPSPIKPAQCKTTA-S	254	
Db	198	QMTBOTCRVLLIGSGOTVGAPASPELPESTGAASSRKPLPLPSPIKPAQCKTTA-S	257	
QY	255	SSSSSSSGAEPKQAFPHLYHNPISGSHPFSGVGGVGRSHKPTPAFSPALPGSTQDLASP	314	
Db	258	SSSSSSSGAEPKQAFPHLYHNPISGSHPFSGVGGVGRSHKPTPAFSPALPGSTQDLASP	317	
QY	315	HLAFPGTGLLAAQCSGANGLEAASBGLLPRKNGSELGYEYISSLEKPGGRHKCRF	374	
Db	318	HLAFPGTGLLAAQCSGANGLEAASBGLLPRKNGSELGYEYISGEVWGPLEKPGGRHKCRF	377	
QY	375	CAKVGSDSALQHLNLSHTEBERPKCVNGCRFTTNGMLKVFNHNRREKYPHQMNPV	434	
Db	378	CAKVGSDSALQHLNLSHTEBERPKCVNGCRFTTNGMLKVFNHNRREKYPHQMNPV	437	
QY	435	PEHLVYVITSSGLPYGMSVPRPKAEEBAGTGGGVERKPLVASTALSTESTLLSTGT	494	
Db	438	PEHLVYVITSSGLPYGMSVPRPKAEEBAGTGGGVERKPLVASTALSTESTLLSTGT	497	
QY	495	STAVAPGLTFENKFVLMKAVEPKSKADENTPRGSEGSALAGVADSGSATRMOLSKLVTSL	554	
Db	498	GATATPGLTFENKFVLMKAVEPKSKADENTPRGSEGSALAGVADSGSATRMOLSKLVTSL	557	
QY	555	PSWALLTNLKSSTGSPRRFVULBPLASPSSTSKLOOLVEKIDRQGAVANVASGAPTT	614	
Db	558	PSWALLTNLKSSTGSPRRFVULBPLASPSSTSKLOOLVEKIDRQGAVANVASGAPTT	617	
QY	615	SAPABSSSA-SGPNQCVTLCLRVLSGPRALRYLHNGONGGERPKCVGSGRAFTSGNLRH	673	
Db	618	SAPABSSSA-SGPNQCVTLCLRVLSGPRALRYLHNGONGGERPKCVGSGRAFTSGNLRH	677	
QY	674	FVGHKTSPPARAONSCPTCKKFTNAVTLOOHVNRHLSGQIRNGSALSSEGGGAOENSS	733	
Db	678	FVGHKTSPPARAONSCPTCKKFTNAVTLOOHVNRHLSGQIRNGSALSSEGGGAOENSS	737	
QY	734	POSTSGPSPFPOQSQRPQRPDEEMS-EEREDEDEEEDVDDEDLSAGSSEGGKATIS	792	
Db	738	POSTSGPSPFPOQSQRPQRPDEEMS-EEREDEDEEEDVDDEDLSAGSSEGGKATIS	797	
QY	793	VAGDEEVGAAEEVATSVAPATTVEKEMDSNEKARQHLTPRRPPRPNLDNDROPMEQTS	852	
Db	798	VAGDEEVGAAEEVATSVAPATTVEKEMDSNEKATQOSSLRRPPRPNLDNDROPMEQTS	857	
QY	853	DVSGAMEEBAKLEGISSPMALTOEGEGTSTPLVEELNLPEAMKKDPGESSGKACVCG	912	
Db	858	DVSGAMEEBAKLEGISSPMALTOEGEGTSTPLVEELNLPEAMKKDPGESSGKACVCG	917	
QY	913	OSFPPOTALEEHOKTPKQDGLFTVCVSCROGFLRATLKHNLNLHNQVRPAPHQONT	972	
Db	918	OSFPPOTALEEHOKTPKQDGLFTVCVSCROGFLRATLKHNLNLHNQVRPAPHQONT	977	
QY	973	ATLSLVPGCCSSITPSGLSPFPRKDDPTIP	1002	
Db	978	ATLSLVPGCCSSITPSGLSPFPRKDDPTIP	1007	

AL3 HUMAN STANDARD; PRT: 1300 AA.
G99X9; G99X9; 15-JUN-2002 (Rel. 41, Created)
15-JUN-2002 (Rel. 41, Last sequence update)
15-JUN-2002 (Rel. 41, Last annotation update)
Sal-like protein 3 (zinc finger protein SALL3) (hsSALL3).
SALL3.
Homo sapiens (Human).
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NCBI_TaxID=9606;
[1]
SEQUENCE FROM N.A. (ISOFORM 3).
Gooding R., Angelicheva D., Blechschmidt K., Svoboda K., Molnar M.,
Tournev I., Kalaydjieva L.;
"Exclusion of HSAL3 and refinement of the region for the CCFDN
gene.";
Submitted (FEB-2001) to the EMBL/Genbank/DBJ databases.
[2]
SEQUENCE OF 29-1300 FROM N.A., AND ALTERNATIVE SPLICING.
MEDLINE=20079154; PubMed=10610715;
Kohlhase J., Hausmann S., Stojmenovic G., Dickens C., Bink K.,
Schulz-Schaeffer W., Altman M., Engel W.;
"SALL3, a new member of the human spalt-like gene family, maps to
18q23.";
Genomics 62:216-222(1999).
-1- FUNCTION: Probable transcription factor.
-1- SUBCELLULAR LOCATION: Nuclear (Probable).
-1- ALTERNATIVE PRODUCTS: 3 isoforms; 1, 2, 3 (shown here) and 4; are
produced by alternative splicing. Isoforms 1 and 2 lack two zinc
finger domains and are the major isoforms.
-1- TISSUE SPECIFICITY: Widely expressed in adult with highest levels
in heart. Expressed in fetal brain (in neurons of hippocampus,
cortex, mediodorsal and ventrolateral thalamic nuclei, putamen,
cerebellum and brainstem).
-1- DEVELOPMENTAL STAGE: In fetal brain of the 24th gestational week.
-1- SIMILARITY: BELONGS TO THE SALL FAMILY OF C2H2-TYPE ZINC-FINGER
PROTEINS.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
at the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; AF347021; AAK18311.1; -
EMBL; AJ007421; CAB65124.1; -
HSSP; P07248; IARE.
GeneW: HGNC:10527; SALL3.
MIM: 605079; -
InterPro: IPR000822; znf_C2H2.
Pfam: PF00096; zf_C2H2; 10.
SMART: SM00355; znf_C2H2; 10.
PROSITE: PS00028; ZINC_FINGER_C2H2_1; 9.
PROSITE: PS50157; ZINC_FINGER_C2H2_2; 8.
Transcription regulation; DNA-binding; Nuclear protein; Zinc-finger;
Metal-binding; Repeat; Alternate splicing.
Zn_FING 420 442 C2H2-TYPE.
Zn_FING 448 470 C2H2-TYPE.
Zn_FING 679 701 C2H2-TYPE.
Zn_FING 707 729 C2H2-TYPE.
Zn_FING 739 761 C2H2-TYPE.
Zn_FING 977 999 C2H2-TYPE.
Zn_FING 1005 1027 C2H2-TYPE.
Zn_FING 1113 1135 C2H2-TYPE.
Zn_FING 1141 1163 C2H2-TYPE.
DOMAIN 198 201 POLY-ALA.
DOMAIN 213 216 POLY-GLN.
DOMAIN 897 902 POLY-SER.
VARSPPLIC 1 133 MISSING (IN ISOFORM 2 AND ISOFORM 4).

FT VARSPPLIC 973 1044 MISSING (IN ISOFORM 1 AND ISOFORM 2).
FT CONFLICT 235 235 R -> C (IN REF. 2).
FT CONFLICT 593 593 V -> L (IN REF. 2).
FT CONFLICT 787 787 D -> N (IN REF. 2).
FT CONFLICT 797 802 DDMDE -> NDNLDK (IN REF. 2).
FT CONFLICT 808 808 D -> N (IN REF. 2).
FT CONFLICT 1138 1138 F -> K (IN REF. 2).
FT CONFLICT 1141 1141 F -> S (IN REF. 2).
SQ SEQUENCE 1300 AA; 135371 MW; 82CF3BDCB8D59150 CRC64;
Query Match 25.2%; Score 1338; DB 1; Length 1300;
Best local similarity 32.2%; Pred. No. 6.2e-51;
Matches 377; Conservative 117; Mismatches 342; Indels 334; Gaps 40;
Db 4 ETGSSRLGGPCGPAPRERGDASEHHHPQVCAKCAQPSDFPEFLAHQNSCCDPPPMVI 63
38 DSGPESRSGG-----EETSVKCCAEFFKMADFLEHQSCTKLPLVLI 82
64 ICGQNPSSSASAPRPEGHSQVMDTEHNSNPDSGSGPPDPWGERGESESGQF 123
83 -----HEDAPAP-----RHEDPE-----PSPASSSERASEAE- 113
Qy 124 LVAATGAAGGGGLLASE-----KLQATPLPESTPPAPPPPPPPGVSGHLN- 176
114 -----EAGAEAGEGEARPEVEKEAEPMDEAPGDTTRAPRPPAPAPPTPAYGASTNV 166
Qy 177 -----IPILFELRVLCQRIHOMQMEQI 201
Db 167 TLEALLSTRVAAVAFSOGARAGGSGAGGVAANVPLLEDGLMALQOQDIDLDLEQI 226
Qy 202 CROVLLLG-----SLQTVGAP-SPSELPGTGAASSTRPLPLFSPIKPAOTGKT 252
Db 227 RSQVALMQRRPPRPSLSAPAPASAPAPAPAPAPAPAPAPAPAPAPAPAPAPAP 266
Qy 253 ASSSSSSSSGAPPPQAEFLHYHPL-----GSOHPFVGVGRSHKPPAPAPALP- 304
Db 267 SAGAPAAALAGSGPAPAPAEAGAPLSPRESGASTPPGAPAESAPAPAPAPAPAPAP 326
Qy 305 -----GSTDQLIASPHLAFPGTGLLAQCIGARGIEAASP----- 342
Db 327 PAPQSAASPOASATPPALAPGS-----LLGAPGLPSPLLPOTASGVTFPPNLYS 379
Qy 343 -----GLKPKNKS-GEIYGGEVYSLEKPGGRHKCRFCAYVGSALQIHL 389
Db 380 IATFANALDPLSLMKHRRKKPRPVSVFEPRKASADPFKHKCRFCAYVGSALQIHL 439
Qy 390 RSHTEBRPKYKCNVGNRFTTGNLKVHFRHREKYPHYOMNHPVNEHLDYVITSSGLPY 449
Db 440 RSHTEBRPKYKCNVGNRFTTGNLKVHFRHREKYPHYOMNHPVNEHLDYVITSSGLPY 499
Qy 450 GMSVPPER-----AEEAGTGGGVGRKPLVA 476
Db 500 GMSLPPERKVTYTLDSKPVLPYPTSVGLQLEPPYVGAHGYADSPATRASRPQRSPRA 559
Qy 477 STTALATFESTLLSTGTS-TAAVAP-----GLPTNKFFVLMKAVEPKSKADENTPPGSEGS 531
Db 560 SSECASLSPLGLNVEGVSATAPESQSLGGRPVYK-----AEVYSLPOTNARAD--- 610
Qy 532 AIAGYADSSATRMQSLKVLTSPLSMALLTNHLKSTGSPFFPYVLEPLGASPESTSLQ 591
Db 611 --APVGAQASAPRTSVGAPTSLSGSLPAPVSEQFAKPFPGFLDLSM--QTSETSKLQ 666
Qy 592 LVEKIDROGAVANAVASAGAPPTTSAPAPSSASGNOVCIVLCRVLSCPALRLHYGHHG 651
Db 667 LVENIDRK-----MTDPNOCYICRVLSQSAKMKHYRTHTG 703
Qy 652 ERPFCKYCGRAFSTRGNLRAHFVGHKTSAPAARQNSCPICOKKFTNAVTLQOHVRLHG 711
Db 704 ERPFCKYCGRAFSTRGNLRAHFVGHKTSAPAARQNSCPICOKKFTNAVTLQOHVRLHG 763
Qy 712 GOIPNGSALSBGGAAQEN--SSEOSTASGPGSFPPOSQOSPSEENSEEBDEDEEE 769
Db 764 GOIPN--FPLPEGFQAMDSLAYDKNAETLSY-----DDMDENSMEDDAEL 811

QY 770 EDVTD-----SLAG-----RSESGEK 789
 Db 812 KDAATDPKPLLSYASCPSPSVISSIALENOMKIDSVSCQOLTLGKSVENGSGE 871
 QY 790 AIVSGDSEEVGAEEVATVAPPTVKENDSNEKAPQHTLPPPPPPDNLDPQPMQ 849
 Db 872 SDRLNSDSSAVGDLE--SRKSGSPAL-----SESSSQALSPANSNGESFRSKSP--- 920
 QY 850 GTSVDSGAMEEAKLEGISSPMALNDEGECTSPPIVE-----LNLPEAMKKDPESSG 904
 Db 921 GUGAEEQEQEILKTERPPSPAAPGSGAPGRAGIKAEADPFLFLSREKGCPS---- 976
 QY 905 RKACEVCGSPPTOTALERHOKTHPKDGLFTCYFCROGFLDRATLKKMLAHN--OYV 962
 Db 977 -TVCCVCGKPFACKSALETHYRSHTERP-FVICALCRGCGSTMGWLKQH-LITHRLKEIP 1033
 QY 963 P--FAPH---GQNTATLELVPCSSSIPS 987
 Db 1034 SQLFDPNFPALGPSQ-STPSLI---SSAAPT 1059
 RESULT 4
 SALI_HUMAN
 AC Q9NSC2; Q9NSC3; Q99881; Q9P1R0; PRT; 1324 AA.
 DT 15-JUN-2002 (Rel. 41, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Sal-like protein 1 (Zinc finger protein SAL1) (Spalt-1like transcription factor 1) (Hsall1).
 OS SAL1 OR SAL1.
 OC Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 NCBI_TaxId=9606;
 [1]
 SFOUNCE FROM N.A., DISEASE, AND VARIANTS SER-150 DEL; SER-150 INS AND G1R-159.
 RX MEDLINE=99138656; PubMed=9973281;
 RA Kohlase J., Taschner P.E.M., Burfeind P., Pasche B., Newman B., Balanc C., Breuning M.H., ten Kate L.P., Maaswinkel-Mooy P., Mitulla B., Seidel J., Kirpatrick S.J., Pauli R.M., Margowski D.S., Devriendt K., Proesmans W., Gabrielli O., Coppa G.V., Mesby-van Sway E., Trembach R.C., Schinzel A.A., Reardon W., Seemanova E., Engel W.;
 RA Molecular analysis of SAL1 mutations in Townes-Brooks Syndrome.;
 RA Am. J. Hum. Genet. 64:435-445(1999).
 [2]
 SFOUNCE OF 1-26 FROM N.A., DISEASE, AND VARIANTS SER-164 DEL AND G1U-1265.
 RX MEDLINE=20004537; PubMed=10533063;
 RA Marlin S., Blanchard S., Lacombe D., Denoyelle F., Alessandri J.-L., Calzolari E., Drouin-Garrard V., Ferraz F.G., Fourminteraux A., Philip N., Toubiane J.E., Petit C.;
 RA "Townes-Brooks syndrome: detection of a SAL1 mutation hot spot and evidence for a position effect in one patient.";
 RA Hum. Mutat. 14:377-386(1999).
 [3]
 SFOUNCE OF 26-1324 FROM N.A.
 RX MEDLINE=97131507; PubMed=8975705;
 RA Kohlase J., Schuh R., Dove G., Kuehnlein R.P., Jaekle H., Schroeder B., Schulz-Schaefer W., Kretzschmar H.A., Kehler A., Mueller U., Raab-Vetter W., Burkhardt E., Engel W., Stick R.;
 RA "Isolation, characterization, and organ-specific expression of two novel human zinc finger genes related to the Drosophila gene spalt.";
 RA Genomics 38:291-298(1996).
 [4]
 SFOUNCE OF 313-345 FROM N.A., AND DISEASE.
 RX MEDLINE=98065876; PubMed=9425907;
 RA Kohlase J., Wäschermann A., Reichenbach H., Froster U., Engel W.;
 RA "Mutations in the SAL1 putative transcription factor gene cause Townes-Brooks syndrome.";

RL Nat. Genet. 18:82-83(1998).
 RN [5]
 RX DISEASE:
 RX MEDLINE=20381976; PubMed=10928856;
 RA Engels S., Kohlase J., McGaughran J.;
 RA "A SAL1 mutation causes a branchio-oto-renal syndrome-like phenotype";
 RA J. Med. Genet. 37:458-460(2000).
 RL
 CC -I- FUNCTION: Transcriptional repressor involved in organogenesis (By similarity).
 CC
 CC -I- SUBUNIT: Interacts with HDAC1, HDAC2, RBBP4, RBP7, MTA1 and MTA2 (By similarity). Probably associates with Nurd histone deacetylase complex (HDAC).
 CC
 CC -I- SUBCELLULAR LOCATION: Nuclear (By similarity).
 CC
 CC -I- TISSUE SPECIFICITY: Highest levels in kidney. Lower levels in adult brain (enriched in corpus callosum, lower expression in substantia nigra) and liver.
 CC
 CC -I- DEVELOPMENTAL STAGE: In fetal brain exclusively in neurons of the subependymal region of hypothalamus lateral to the third ventricle.
 CC
 CC -I- DISEASE: Defects in SAL1 are the cause of Townes-Brooks syndrome (TBS); a rare, autosomal dominant malformation syndrome with a combination of imperforate anus, triphalangeal and supernumerary thumbs, malformed ears and sensorineural hearing loss.
 CC
 CC -I- DISEASE: Defects in SAL1 are in rare cases associated with hemifacial microsomia (HFM) or Goldenhar syndrome; a common birth defect involving first and second branchial arch derivatives. The highly variable phenotype shows in addition to craniofacial anomalies cardiac, vertebral, and central nervous system defects.
 CC
 CC -I- DISEASE: Defects in SAL1 are associated with branchio-oto-renal (BOR) syndrome, an autosomal dominant disorder manifested by various combinations of preauricular pits, branchial fistulae or cysts, lacrimal duct stenosis, hearing loss, structural defects of the outer, middle, or inner ear, and renal dysplasia. Associated defects include asthenic habitus, long narrow facies, constricted palate, deep overbite, and myopia. Hearing loss may be due to Mondini type cochlear defect and stapes fixation.
 CC
 CC -I- SIMILARITY: BELONGS TO THE SAL FAMILY OF C2H2-TYPE ZINC-FINGER PROTEINS.
 CC
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: Y18265; CAB41400.1; -
 CC EMBL: Y18264; CAB41399.1; -
 CC EMBL: X98833; CAB41399.1; JOINED.
 CC EMBL: AF017655; AAB99908.1; -
 CC EMBL: AF074949; AAF19263.1; -
 CC HSSP: P15822; 1BPO.
 CC GeneW: HGNC:10524; SAL1.
 CC MIM: 602218; -
 CC MIM: 107480; -
 CC MIM: 164210; -
 CC MIM: 113650; -
 CC InterPro: IPR000822; Znf_C2H2.
 CC Pfam: PF00096; Zf-C2H2; 10.
 CC ProDom: PD000003; Znf-C2H2; 1.
 CC PROSITE: PS00026; ZINC_FINGER_C2H2_1; 9.
 CC PROSITE: PS50157; ZINC_FINGER_C2H2_2; 9.
 KW Transcription regulation; Repressor; DNA-binding; Nuclear protein;
 KW Zinc finger; Metal-binding; Repeat; Polymorphism.
 FT ZN_FING 449 471 C2H2-TYPE.
 FT ZN_FING 477 499 C2H2-TYPE.
 FT ZN_FING 706 728 C2H2-TYPE.
 FT ZN_FING 734 756 C2H2-TYPE.
 FT ZN_FING 766 788 C2H2-TYPE.
 FT ZN_FING 1001 1023 C2H2-TYPE.
 FT ZN_FING 1029 1051 C2H2-TYPE.

2N_FING 1134 1156 C2H2-TYPE.
 2N_FING 1162 1184 C2H2-TYPE.
 DOMAIN 150 159 POLY-SER.
 DOMAIN 160 163 POLY-SER.
 DOMAIN 237 240 POLY-SER.
 DOMAIN 294 297 POLY-SER.
 DOMAIN 371 375 POLY-SER.
 DOMAIN 1144 1147 POLY-SER.
 VARIANT 150 150 /FTID=VAR_013155.
 VARIANT 150 150 /FTID=VAR_013156.
 VARIANT 159 159 /FTID=VAR_013157.
 VARIANT 164 164 /FTID=VAR_013158.
 VARIANT 1265 1265 /FTID=VAR_013159.
 CONFLICT 79 79 G -> A (IN REF. 1: CAB41399).
 SEQUENCE 1324 AA: 140390 MW: CBE787847A96D8AA CRC64;

Query Match 24.5%; Score 1301; DB 1; Length 1324;
 Best Local Similarity 31.3%; Pred. No. 2.5e+49;
 Matches 383; Conservative 142; Mismatches 338; Indels 360; Gaps 48;

4 ETGSSRLGPGC--GEPAERGGDAEEHHROYCAKCAQFSDPTERFLAHQNSCTDPPVM 61
 17 EVASLPKRRHGDTEKGGPS-RPTKSKDAH--VCGRCAEFFELSDLLHKKNTCKNQLVL 72
 62 VIIGGOENPSNSASAPRP-----EGHSRSQVMDTEH 94
 73 IV---NENGSPETSPSPRPNDPQMDNTVNTKTDQVDCDLSLHNGDRESMEVEA 129
 95 --SNDDSGSSGPPDPTWGPERRGESSGOFLVAATGTAAGGGGCLILASPKEATPLPP 152
 130 PVANKSGSGTSS-----GSHSTAPSSSSSSSGSGSSSTGTSATITSLPQ 178
 153 -----ESTP-APPPPPPPPPPPVSGSHINIPILIEELRYLQO 189
 179 LGDLTLTGNFVINSVNIINLSTKYAVAOFSQBARCGASGKLAVPALMEQLLALQO 238
 190 ROIHOMOMTEOICROYLLGLSGTVGAPASPSLELPGT----- 227
 239 QOIHQLOLEQIHNHQLLLAS--QNALDPTSSPSGQTLRTSANPLSTLSSHLSQLAA 296
 228 -----GAASSTKPLPLFSP-----IKPAQTK-----TTASSSS 258
 297 AGIAGSLASQASASISGVKQLPPIQLPOSSSGNTIIPNSGSSPFNMNIIAAAVTPPSSEKV 356
 259 SSSSGAEPKQAFHLYHPL---GSQHPFVGVGRSHKPTPAPSPALP--GSTDQLIAS 313
 357 ASSAGAS-----HNSNPVSSSSSPAFAT-----SLLSPASNPILLPOOASANSVPPS 404
 314 PHLAEPGTGLLAOCIGAAAGLEAAASPLGLKPKNGSGELGYEYISLEKPGCRHKCR 373
 405 P-LPNIQT---AEDLNS---LSALAOQRKSKPNNVTA---FEAKSTDEAFKHCCR 452
 374 FCACVFGSALQIHLSSHGERPYKVCNGRPTTRGNLKVHFHRRKEXPHVOMNHP 433
 453 FCACVFGSALQIHLSSHGERPYKVCNGRPTTRGNLKVHFHRRKEXPHVOMNHP 512
 434 VPEHLDYITSSGLPYGMSVPERKAEEAGTPGGVERKPLVASTALASSTLSTG 493
 513 VPEHLDNIPSTGIPYMSIPPE-----KPV---TSLMDKPVLPPLTTS 554
 494 TSTAVALPGLTFEKFLAKAVEPKS---KADENTPPGSEGSATAGVADSG---SATR--- 544
 555 VGLPLPPLSLPLPFI--KTEEPAPIPISHSATSPGYSKS-----DSGGPEASATNMG 606
 545 -----MOLSKIYV-SLP-----SMALLTNTL----- 564
 607 GLPEAEGSTLPPSGGKSEESGAVTNSVPTASSSVLSPADCGPAGSATITFTPLFLM 666

QY 565 --KSTGEPPPYLEPLGASPSSETSKLOOLVEKIDROGAVAVASTAGAPTTSAPAPSS 622
 Db 667 SEQKAKAFPEFGGLD--SAQASETSKLOOLEVENIDKK----- 701
 QY 623 ASGNOCVICTLRVISCPRALRLHYGOGBERPKCKVCGARAFSTRGNLRAHFVGKTSRA 682
 Db 702 ATPNECITCHARVLSQOSAKLMHYRTHGERPKCKICGRAFTTKGNKTHYSVRAMP 761
 QY 683 ARANOSCPICQKFTNATVLOQHVRHNLGGQIPN-----GGSALSGGAOEN----- 731
 Db 762 LRYOHSOPIQCKFTNNAVYLQQHIRMHMGQIPNTPVPDSTSESMSDTSDEKKNFDL 821
 QY 732 ---SSEOSTAGSPGSPQ--POGQOPSPPE-----EMSE----- 760
 Db 822 DNFEDENMEDPEGSIPDTPKADASQDLSSSBPLPLEMSIALLENQMKMINAGLAEOL 881
 QY 761 ---EEEDDEEEDVDTDSDLAGRSESGEKAISYRGDSSEVSGAEEVATVAAPT 816
 Db 882 QASLSYVNGSIEGDVLTNDS-----SSVGGMESQAGSPAPISESTSSMOA 928
 QY 817 VKEMDSNF---KAPQHTLPPPPPPNDLH---PQPMEOGTSVSGAMEEAKLEGISSP 870
 Db 929 LSPNSIQEFPKSPSIEEKQRAVPSEFANGLSFTPVNGGLDLTSSRAK----- 979
 QY 871 MAALQEGEGTSTPLVEELNLEPMARKDPGESSGRKACEVCGQSPPTOTALSEHOKTHPK 930
 Db 980 --IKEDSLGILFP-----RDGKFK-NFACDICKGTFCACQALDIHYRSHK 1025
 QY 931 DGPLTFCFCROGFLDRALTKKMLAHQV-----PPAPPGQNIATLSLVPGS 982
 Db 1026 ERP-FICTVNCNRFSTKGNLKKOMLTL--HOMRDLPSOLEFEPSSNLGPNONSAVTPANSL 1082
 QY 983 SSIPSP-----GLSPFRKDDPT 1000
 Db 1083 SLITKVGPFVHVSPOQSKDTP 1105

RESULT 5
 SAL3_MOUSE
 ID SAL3_MOUSE STANDARD: PRT: 1323 AA.
 AC Q62255;
 DT 15-JUN-2002 (Rel. 41, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Sal1-like protein 3 (Spalt-like protein 3) (Msal) (Fragment).
 GN SAL3 OR SAL.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain, and Embryo;
 RX MEDLINE=96391179; PubMed=8798152;
 RA Ott T., Kaestner K.H., Monaghan A.P., Schuetz G.;
 RT "The mouse homolog of the region specific homeotic gene spalt of Drosophila is expressed in the developing nervous system and in mesoderm-derived structures.";
 RL Mech. Dev. 56:117-128(1996).
 CC -1- FUNCTION: Probable transcription factor.
 CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).
 CC -1- ALTERNATIVE PRODUCTS: 2 isoforms; 1 (shown here) and 2: are produced by alternative splicing. Isoform 2 lacks two zinc finger domains (6 and 7) and is the major isoform.
 CC -1- TISSUE SPECIFICITY: In adult brain, testis and kidney. In lower levels also in adult ovaries and embryonic stem cells. In embryo in developing neuroectoderm of brain, inner ear and spinal chord. Also weakly and transiently expressed in embryonic branchial arches, notochord, limb buds and heart.
 CC -1- DEVELOPMENTAL STAGE: During embryogenesis detected from 7 dpc onward in tissues derived from mesoderm and ectoderm.
 CC -1- SIMILARITY: BELONGS TO THE SAL FAMILY OF C2H2-TYPE ZINC-FINGER PROTEINS.

CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: X97581; CAA66196.1; -
DR HSSP: P07246; IARE.
DR MGD: MGI:109295; Sa113.
DR InterPro: IPR000822; Znf.C2H2.
DR Pfam: PF00096; zf-C2H2; 10.
DR SMART: SM00355; Znf.C2H2; 9.
DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; 9.
DR PROSITE: PS50157; ZINC_FINGER_C2H2_2; 8.
KW Transcription regulation; DNA-binding; Nuclear protein; Zinc-finger;
KW Metal-binding; Repeat; Alternative splicing.
FT NON_TER 1
FT ZN_FING 398 420 C2H2-TYPE.
FT ZN_FING 426 448 C2H2-TYPE.
FT ZN_FING 663 685 C2H2-TYPE.
FT ZN_FING 691 713 C2H2-TYPE.
FT ZN_FING 723 745 C2H2-TYPE.
FT ZN_FING 968 990 C2H2-TYPE.
FT ZN_FING 996 1018 C2H2-TYPE.
FT ZN_FING 1104 1126 C2H2-TYPE.
FT ZN_FING 1132 1154 C2H2-TYPE.
FT DOMAIN 119 123 POLY-PRO.
FT DOMAIN 188 191 POLY-GLN.
FT DOMAIN 345 348 POLY-SER.
FT DOMAIN 881 886 POLY-SER.
FT VARSPIC 964 1035 MISSING (IN ISOFORM 2).
FT SEQUENCE 1323 AA; 139070 MW; 86D83EDEAF2CEDD0 CRC64;
Query Match 24.2%; Score 1282; DB 1; Length 1323;
Best Local Similarity 31.3%; Pred. No. 1.6e-48;
Matches 374; Conservative 133; Mismatches 329; Indels 360; Gaps 45;
4 ETGSSRLGCGCGEARECGDASEHNHPOVCAACCAQFSDTEFLAHONSCCTDPPVMT 63
D 9 DSGSESRG-----SEE--TSVCEKCCAEFFKWADELOHKTKKNPLVLIV 53
QY 64 IG-----GQENPSSSASS-APREGHRSROYMDTEHS-----NPPDSSSGSP 105
D 54 HDDEPARPSEDPPEPSPASSPSDRTSEVAEEVAAPTEGSEVKAATKEASMDVEVSTDKP 113
QY 106 PDPPTWGPERRGEESGQFLVAATGTAAAGGGGLILASPKLGATPLPESTPAPPPPPP 165
D 114 PGP-----SVPPPPALPQQ 128
QY 166 PPP-----GVGS--GHNLIPLLEELRVLQ 188
D 129 PEPAFASMPSTNVTLETLSTVVAVAQFSQAGARAGTTGAGSVAVALPMLLEQLVALQ 188
QY 189 QROIHOMQMTBOICHQVLLGSLGTVGAPASPSL-----P 225
D 189 QQQHQHQLDLEQIRQV---GLMSKROPGLPKASAPGNTSVOLGTLPHNALQLSACP 245
QY 226 GTGAASSTKPLPLPFSPIKPAQTGKTTASSSSSSSGAEPPKQAFHLYPLIGSQHP-- 283
D 246 ATASAGSGSTLPAPAD--GPQHLSPASGTSSTPCCSAAPRDSG---AHPACSTGPAP 299
QY 284 ----FSVGVGGRSHNP-----TP-----APSPALPESTDOLIASPHLAPGT 321
D 300 GAVAAASSTVGAAVPOQNASTPALGPGPLLSASASNLPPNLLPQT-----SSSSVTFPP 354
QY 322 TGLLAAGCLGARGLEAAASPGILKPKNGS-GEIGYGEVVISLEKPGGRHRCRFCAKVG 380
D 355 LVSTIAA-----TANALDPLSA--LMKHRRKGRPNVSVPEPKASADPPFKHKRCRFCAKVG 408
QY 381 SDSALQIHLRSHTGERRPKCNVCGNRFRTTGNLKVHFRHREKYPHYOMNPHVPEHLDY 440

|||||
Db SDSALQIHLRSHTGERRPKCNVCGNRFRTTGNLKVHFRHREKYPHYOMNPHVPEHLDK 468
QY 441 VITSSGLPYGMSVPERK----AEEAGTP---GGVERKPLVASTALSATSTLTL 491
D 469 CPTCSGIPYGMSPILRCPVTTWLDSPVLPLVPTSVGLDPLPVTGTHYTDSPSTIPVS 528
QY 492 -----TGSTAAVAPGLPTFNKPFVLMKAVERK-----SKADETPPGSE---GS 531
D 529 RSPQRPSPASSSECTSLSPGINTNESGITVRESPOPLLGPSLTAAEPVSLPCTSTRTGD 568
QY 532 A-IAGVADSG-----SATRMQLSKLVTSL--PSMALITNLKSTGSPFPVPLPLGASPS 585
D 569 APVYGQVSGLPSTASATPAVYDSACTSLGSPGLPRAVSDPFA--QPPFGLLDSM--QTS 644
QY 586 TSKLDQLEVKIDROGAVAVASTAGADPTTAPADSSASGPNQVICYLVLSCPRALRH 645
D 645 TSKLDQLEVENIDK-----MTDENQVICYHVLSCQSALKMH 681
QY 646 YGQHGGERPFCKKVCGRAPSTRGNLRAHFVGHKTSPPARAQNSCTICQKFTNAVYLOOH 705
D 682 YRTHGERPFCKKICGRAFTTKGNLKTGHGVHGRKPLRVGHSCPTCQKFTNAVYLOOH 741
QY 706 VAMHLGGQIPN-----GGSALSEGGAQDENSSPO-----ST 737
D 742 IRMHNGQIIPNTPPLPGLORAMADLPFDEKNAETLSSFDIDENSMEDSEIKRTASD 801
QY 738 ASGP-----GSFPQPOSQDPSPEEMSE-----BEEDEEBE 769
D 802 SSKPLSTYSGSCP-----PSPSVISSIALLENQMKMIDSVNMQCOLANLKSVENSGE 855
QY 770 EDVTEDESLAGRS---SESGEKAISVRGDSSEVSGAEEVATVAAPTTVEMDSNEKA 826
D 856 SDRLSDSSAVGDLESRSAGSPALSSSSQALSPANSGBEFSRKSGLHQEPQET 915
QY 827 POHT--LPPPPPPDNLDHPQPMEOGTSDVSGAMEEAKLEGISSPMALTOEGEGSTP 884
D 916 PLKTERLSDPPPGPN-----GGALDLTGTGNGP-----RP 945
QY 885 LVEELNLPPA-----MKDPGESSGKACVCGQSFPTOTALAEHQTKHPDGLPTCV 938
D 946 LIKE-----EAPSLILFLSRGRKCAAS-TVCGVCGKPFACKSALLETIHYHSHTKE--RFFVCT 999
QY 939 FCRQGFELDRATLKKHMLAHNOVPPFAPHGQNTATLSLVGCCSSISPSGLSPFP 994
D 1000 VCRRCSTMGNKQHLTL--HKLKEI---PSQVFDPPFTLGPSHTSDSLASSPAP 1049
RESULT 6
SALL_MOUSE
ID SALL_MOUSE STANDARD; PRT; 1322 AA.
AC Q9ER74; Q920R5;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Sal-like protein 1 (Zinc finger protein Spalt-3) (Sal-3) (Msal-3).
GN SALL1 OR SAL3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniala; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_Taxid=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=129/Ola;
RX MEDLINE=20422467; PubMed=10965108;
RA Buck A., Archangelo L., Dixkens C., Kohlase J.;
RT "Molecular cloning, chromosomal localization, and expression of the
RL murine SALL1 ortholog sal1-1.";
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=21543523; PubMed=11688560;
RA Nishinakamura R., Matsumoto Y., Nakao K., Nakamura K., Sato A.,


```

Db      73 IV--NESPASPKATFPBPBSLNDPPDDMKRAANKADQEDCSDLSEPK-GLDR--EESME 126
QY      122 QRLVATGT-----AAGGGGLIASKLGATP-----LPP----- 152
Db      127 VEPVATTTTTTTGGSSGSGSTLSGVNTITPFSCHSCSSSTSAITTSLPQGLDTLGG 186
QY      153 -----ESTP-APPPPPPPPPGVSGLNPLILEELRYLQORQHOMQ 197
Db      187 NFSVINSNVITENIQSTKRVAVACPSOEARCGASAGSKLLISTLMOLLALQOOQIHQLOL 246
QY      198 TEOICROVILLIGLQGVGAPASPSSELPCTGA-ASTPRLPLFEPIK---PAQTKTTA 253
Db      247 IEOIRHOLLIAS--QSADIPAAIS-IPSGTLKRSANPLITLSSHLSQULAVAAALQAS 303
QY      254 SSSSSSSSSGAE-----PPKQAFHILYHPLGSOH-FSV-----GGYGR 291
Db      304 LASOSANISGVKOLPHVOLPOSSSGTSTIVPSSGTSPPMKSITVTAAVPTPPSSSEKVASMAQA 353
QY      292 SIKPTPA-----PSPA-----LPSTODLISP---HLAFPTGTGLLA--AOCILGAARGL 336
Db      364 SHVSSRAVASASSSPAFKAISSLSBESNLLRQPTPAAVFPRLPLNIAATTAEDLNS---L 420
QY      337 EAAASPELLKPKNGSELGYGEVITSLKPCGRBKRCRCACAKYFGSDSALOIHRSITGER 386
Db      421 SALAAOORKSRRPNTVA---FEAKSTSDAEAFKRCRCACAKYFGSDSALOIHRSITGER 476
QY      397 PYKCANVCNGFTTRGNLKVHNRHREKRPVHOMNHPREHNDVYTSSGGLPYGMSVPRE 456
Db      477 PFCNIGNCFSTKGNLKVHQRKKEKRPHTIOMNPRVREHNDVPTSTGIPYMSIPSE 536
QY      457 KAEELAGTPEGGVYERKRLVASTTSLAESLSTLSTGSTVNAVAGLPPTNKKFLMAVPE 516

```

[illegible]

```

OY 844 --POPMOGISDVSGAMEEAKLEISSPMALTOBEGTSPYBELNPEAMKDDPE 961
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 959 LSPFPVAGGALDLTSSNAEK-----IKRDSLGILPFP-----RDRCK 996

OY 902 SSGRKACEVCGSEFPYOTALAEHOKTHPKDGLFTFCVCRGEGFLDRATIKHMLAHQY 961
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 997 FK-NTACDIDCGKTTACOSALDIHYRSHKRP-FICTYCNRGSTGKNLQKHILT--HQM 1053

OY 962 -----PEFAPHPQNIATLILVPGCCSSIPSP-----GLSPFRKDDPT 1000
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 1053 RDLPSQLFEFSSNIGPROMNSAVIPRANSSLLIKTEVNGGVNHSPOSKDAFT 1104

```


RESULT 7
ID SALM_DROME STANDARD; PRT: 1355 AA.
AC P39770: 15-FEB-1995 (Rel. 31, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Homeotic protein spalt-major.
SALM.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxId=7227;
[1]
RP SEQUENCE FROM N.A.
RA MEDLINE=94139659; PubMed=7905822;
RA Kuehnlein R.P., Frommer G., Friedrich M., Gonzalez-Gaitan M.,
Weber A., Wagner-Bertholz J.F., Gehring W.J., Jaechle H., Schuh R.;
RT "Spalt encodes an evolutionarily conserved zinc finger protein of
RT novel structure which provides homeotic gene function in the head and
RT tail region of the Drosophila embryo.";
RL EMBL J. 13:168-179(1994).
CC -1- FUNCTION: REQUIRED FOR THE ESTABLISHMENT OF THE POSTERIOR-MOST
CC HEAD AND THE ANTERIOR-MOST TAIL SEGMENTS OF THE EMBRYO. PROBABLY
CC FUNCTION AS A TRANSCRIPTIONAL REGULATOR. COULD REPRESS THE
CC TRANSCRIPTION OF THE TSH GENE.
CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -1- DEVELOPMENTAL STAGE: FIRST EXPRESSED AT BLASTODERM STAGE AND
CC LATER IN RESTRICTED AREAS OF THE EMBRYONIC NERVOUS SYSTEM AS
CC WELL AS IN THE DEVELOPING TRACHEA.
CC -1- SIMILARITY: BELONGS TO THE SAL FAMILY OF C2H2-TYPE ZINC-FINGER
CC PROTEINS.
CC
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: X75541; CAA53229.1; -
DR HSP: P15822; IABO.
DR FlyBase: FBgn004579; salm.
DR InterPro: IPR000822; Znf_C2H2.
DR Pfam: PF00096; Zf-C2H2; 7.
DR PRINTS: PR00048; ZINCFINGER.
DR SMART: SM00355; Znf_C2H2; 7.
DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; 7.
DR PROSITE: PS00157; ZINC_FINGER_C2H2_2; 7.
KW Developmental protein; Zinc-finger; Metal-binding; DNA-binding;
KW Nuclear protein; Transcription regulation; Repeat.
FT ZN_FING 451 473 C2H2-TYPE.
FT ZN_FING 479 501 C2H2-TYPE.
FT ZN_FING 824 846 C2H2-TYPE.
FT ZN_FING 852 874 C2H2-TYPE.
FT ZN_FING 884 906 C2H2-TYPE.
FT ZN_FING 1289 1311 C2H2-TYPE.
FT ZN_FING 1317 1339 C2H2-TYPE.
SQ SEQUENCE 1355 AA; 148995 MW; 6FC4EBDC0BC6355E CRC64;
Query Match 14.3%; Score 759.5; DB 1: Length 1355;
Best Local Similarity 22.7%; Pred. No. 5.2e-26;
Matches 312; Conservative 141; Mismatches 415; Indels 505; Gaps 44;
3 QFTGSSRLGGPCGEPAERGDASEEHHPVCAKCAQFSDTEFLAHNSCTDPPVMY 62
Db 50 KDIGSDQENGCGSPLTTATTTASPSRSP-----PEEDPEEOSTSQST---PEOST 100
QY 63 IIGGEPNSNSASAPRPGHSHRQVMDTEHSN--PPDSGSSGPPDPTWGPERRGEES 119

Db 101 PDHLENDIKSEAKSEIEPVEDNNNNRYAMTKPSSSEEREPNAGSMSSPV--AEASAE 158
QY 120 SGQFLVAATGTAAAGGGGLLASKLGT-----LPPESTPAPPPPPPP 166
Db 159 ATE-----RTPEKEKEDVEVDVMDPEADSSAVSTEVTL 194
QY 167 PGVSGSHLNPILLELR-----VIG 188
Db 195 PGAGAG-----PYTLEIQNMQALIAQPAATIANGSNGADAMKQLAFLQDTLENLQ 249
QY 189 QROIHOMQTEQICROYLL-----LGSIGQTVGA 217
Db 250 QOOLFQIQLQLOSQALNAQAKOEEDTEADADQDEQDETDTYEERADIAMELRQKA 309
QY 218 PASPSLPGGAASSTPLPLFSPIKPAQIGKTTASSSSSSSSCAEPKQAFPHLYHP 277
Db 310 EARMAE-----AKARQHLINAGVPLR-----ESSGPAESLKRREHDH- 348
QY 278 LGSQHPFVSGVGRSHKPTPA-----PSPALPGSTD--QLTASPH--LAPQTTG 323
Db 349 --ESQPNRRTSLDNTKADTRADALAKLKEMENTPLPGSDLAISITTHNDLPERNSLD 406
QY 324 LLAAGCLGARGLEAASPGILK-----PRNGSELGYEVISSLEKPGRRHCR 373
Db 407 LLOKR-----AQEVLDSASQGLTANSMADDFAFGEKSGECK-----GRNEPFFKHCR 454
QY 374 FCAGVGSDSALQIHLNRSHGGERPKYKVCNCRFTTRGNLAKYHFNHREKRPVYQMPHP 433
Db 455 YGCKVGVSDSALQIHHSHGGERPKVCNCRFTTRGNLAKYHFNHREKRPVYQMPHP 514
QY 434 VPEH-----LDYVITSSQLPYGMSVPERKAEDEAGTGG--GVER--KPLVASTAL 481
Db 515 IPEHMKFHPRLDQMEPTDSSPNHSPAPRPLGSAFPAPSPGIONLYKRPMEILKSL 574
QY 482 SATESLTLLSTGTSTAV--APGLTFNKFYLMKAVERKSKADENTPPSGEGS----- 531
Db 575 GAAAPQVFPQELPTDLRKPSPOLDEDEPOYKNEVEEKDREHQBMAECSEPEREPL 634
QY 532 -AIAGVADSSATRMQSK-----LVTSLPSMALLNHLKSTGFFPPVYLEPL----- 579
Db 635 PLEVRKEEVEDEQVQKQEDHRIEPRTPSPSSSHRSPHNHRSHMGVPPVQPIQPA 694
QY 580 --GASPSSTKLQD-----VEKIDRGAAVAVASTAGAP 612
Db 695 LMRPGSSPGSQSHDLPTREQDLPRREDFAERFLNFTAKMLSEHNSPVASPRAGAL 754
QY 613 TTSAPAP----- 619
Db 755 PRGVPRPHNHPHMASSPFENPIKHEMALLPRHSDNSMNEFTEVNTCETMKLKEL 814
QY 620 ---SSSAGRPQVCYCLVLSCPRALRLHYGONGGERPRCKYCGRAFFSTRGNIRAHVGH 677
Db 815 MKKKKISDPKCCVCDVLSCKSALOMHYRTHTGPRFCRIGRAFTTKGNLKTMAVH 874
QY 678 KTSAPAAQNSCPICQCKFTTAVALVLOOHVHMILG-----GQ 713
Db 875 KIRPRPRNFIQCVCCHKKYSALVLOQHIRLHTGERTDLTPQIOALRIQDPPPMRPGH 934
QY 714 IPNGSALSBGGAAQDENSSEOSTASGSPFPQ-----OSQOPSPPEEMSEEE 762
Db 935 FMPFPAALAAHFGA-----LPGGGPPRPNHGAHNGALGSBSQDMDDNMDCGE 985
QY 763 EEDP-----EEDP-----VTDEDSLGR 781
Db 986 DVDDDVSSSEHLNSNLSQEGDRSRSRDDFKSLLEOKLRIDATGVVNTNPRRSSASSH 1045
QY 782 GSESGEKA---ISVRDSEEV-----SCAEFEVA-----TSVAAPTYYKEMS- 822
Db 1046 GHSVGSTAPTSPSVNHSQYIKRSSSPARSPASGALDLTFRRAATSSSSSSSPLPKK 1105
QY 823 -----NEKAPQHTLPPPPP-----PDNLDH----- 843


```

QY      700 VTLOOHVNRHLG-----GOIPNG---- 717
           : |||::|||
Db       918 LVLOOHIIRLHTGEPTDLTEEOIOAAEIRDPSPMPPGHMPNPFMAAFAHFGAMGGAGG 977
           |
QY       718 -----GSALSEGG----- 725
           |||::|||
Db       978 PRGATGMGPGRPHNGTLGSSSQDDLDDNMDCGDGDDFDIISSEHLSNDPAAITSDRRSS 1037
           |||::|||
QY       726 -----GAQENSSSEQ--STASGPSF-----POQSOPR--- 752
           |||::|||
Db       1038 DDFKSLFECKLRIDPTGVNVINSHQRPHSAASNPSIGSASASPAPSPSSOPKPKSCS 1097
           |
QY       753 -----SPEENSEEEDEEDEEDVTDE--DSLARGSESAGEKAISV----- 793
           |||::|||
Db       1098 PVRSSCSSEVFVS---ETSGDALDILTPRALPRPLASSSRSPYRLQLSVRRRLARSYS 1153
           |||::|||
QY       794 ----- 793
Db       1154 SHRCVPWVALLSOLPPSVGLDCLPPLGIQHNLQQHQNHLMQQQAHAVAANAQAHHNQ 1213
           |
QY       794 -----RGDSBEV--SGAEENVATSVAAPTVMKEMSNEKAQNTLRP--- 833
           |||::|||
Db       1214 MQQNAAALHQHOEHLLREADEVOOKAAQEVOOKAAAAAAAQAQESIQ---PPRSRG 1270
           |
QY       834 -----PPPDPNDLHPQP-----MEGTSDVGAMEEAALKEGISSPMALTQE 877
           |||::|||
Db       1271 ESSVGPPAAPNPFLISA RPPPGMFNP LIFRPATTQNMCMANQIAQSVMAPDAFPNALIS 1330
           |
QY       878 GEGYSTPLVEELNLPBAMMKKDPESSGRKKACEVCGQSFFPTQALIEHKTHKRDGPLFTC 937
           |||::|||
Db       1331 GVRSQT-----TCGICXKTPPCHSALETHYRSHTKERP-FKC 1366
           |||::|||
QY       938 VFCROGFLDRA TLTKRHML 955
           |||::|||
Db       1367 NICDRGFTTKG NLKHOML 1384
           |||::|||

RESULT 9
Z236_HUMAN
AC   O9ULJ6; O9ULJ7; STANDARD; PRT; 1845 AA.
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   zinc finger protein 236.
GN   ZNF236.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX   NCBI_TaxID=9606; [1]
SEQUENCE FROM N.A.
TI   SEQUENCE-Kidney.
MC   MEDLINE=93385731; PubMed=10458916; Holmes D.I.; Wahab N.A.; Mason R.M.; Cloning and characterization of ZNF236, a glucose-regulated kruppel-like zinc-finger gene mapping to human chromosome 18q22-q23." ; Genomics 60:105-109,(1999).
RL   -FUNCTION: MAY FUNCTION AS A TRANSCRIPTION FACTOR.
CC   -SUBCELLULAR LOCATION: Nuclear (Probable).
CC   -ALTERNATIVE PRODUCTS: 2 ISOFORMS: A AND B (SHOWN HERE); ARE PRODUCED BY ALTERNATIVE SPLICING.
CC   -TISSUE SPECIFICITY: UBIQUITOUS. EXPRESSION LEVELS ARE HIGHEST IN SKELETAL MUSCLE AND BRAIN, INTERMEDIATE IN HEART, PANCREAS, AND PLACENTA, AND LOWEST IN KIDNEY LIVER, AND LUNG.
CC   -SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-FINGER PROTEINS.
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial

```

```

CC CC entitles requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF085244; AAD55329.1; -.
DR EMBL; AF085243; AAD55328.1; -.
DR HSSP; P07248; IARD.
DR Genew; HGNC:13028; ZNF236.
DR MIM; 604760; -.
DR InterPro; IPR000822; ZNF_C2H2.
DR Pfam; PF00096; ZF-C2H2; 30.
DR ProDom; PD000003; ZNF_C2H2; 5.
DR SMART; SM00355; ZNF_C2H2; 30.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 30.
DR PROSITE; PS0157; ZINC_FINGER_C2H2_2; 30.
KW Transcription regulation; DNA-binding; Zinc-finger; Metal-binding;
KW Nuclear protein; Repeat; Alternative splicing.
FT DOMAIN 37 1801 ZINC_FINGERS.
FT ZN_FING 37 1801 C2H2-TYPE.
FT ZN_FING 66 88 C2H2-TYPE.
FT ZN_FING 93 115 C2H2-TYPE.
FT ZN_FING 121 143 C2H2-TYPE.
FT ZN_FING 153 175 C2H2-TYPE.
FT ZN_FING 197 219 C2H2-TYPE.
FT ZN_FING 225 247 C2H2-TYPE.
FT ZN_FING 253 276 C2H2-TYPE.
FT ZN_FING 285 308 C2H2-TYPE.
FT ZN_FING 482 504 C2H2-TYPE.
FT ZN_FING 510 532 C2H2-TYPE.
FT ZN_FING 538 560 C2H2-TYPE.
FT ZN_FING 566 588 C2H2-TYPE.
FT ZN_FING 657 679 C2H2-TYPE.
FT ZN_FING 685 707 C2H2-TYPE.
FT ZN_FING 713 735 C2H2-TYPE.
FT ZN_FING 741 763 C2H2-TYPE.
FT ZN_FING 967 989 C2H2-TYPE.
FT ZN_FING 1023 1045 C2H2-TYPE.
FT ZN_FING 1051 1073 C2H2-TYPE.
FT ZN_FING 1166 1189 C2H2-TYPE.
FT ZN_FING 1195 1217 C2H2-TYPE.
FT ZN_FING 1223 1245 C2H2-TYPE.
FT ZN_FING 1251 1273 C2H2-TYPE.
FT ZN_FING 1657 1680 C2H2-TYPE.
FT ZN_FING 1686 1708 C2H2-TYPE.
FT ZN_FING 1722 1744 C2H2-TYPE.
FT ZN_FING 1750 1772 C2H2-TYPE.
FT ZN_FING 1778 1801 C2H2-TYPE.
FT VARSPIC 1550 1558 ELNFTSGSLSTPTTPSPRAISTONLYMS -> GSNVSHS
FT VARSPLIC 1559 1845 VGFQEGSVLALYLENSDKT (IN ISOFORM A).
FT FT MISSING (IN ISOFORM A).
SQ SEQUENCE 1845 AA; 203659 MW; 2879EA91D0C6D3D8 CRC64;

Query Match 7.0%; Score 371.5; DB 1; Length 1845;
Best Local Similarity 20.7%; Pred. No. 3.6e-09;
Matches 241; Conservative 124; Mismatches 45; Indels 345; Gaps 45;

QY 15 CGEPAERGGDASEEH-----HPDYCAKCAQAEFSDPTFLAHONSC-----CT 56
Db 230 CGKANOKG-ALQTHMIKHTGKEPHACAFCPAASQKGNLQSHVORVHSEVKNGPYNTCT 288
QY 57 DPPVH-----VIIGQENSNS-----SASSA 78
Db 289 ECSCVFKSLGSLNTHISKMHGSGPQNSTSTETAHVLTATLFTQLPLQOTEAQTSASSQ 348
QY 79 PRPEGHSR-SQVMDTEHNSNPDSGSGPDPPTGPERRGESGQFLVAATGTAGGGG 136
Db 349 PSSQAVSDVTDQLLELSEAPVSEGS-QPQ-----GQQLSTYGI--NOD 391
QY 137 GLILASPKIGATPLPEESTP-----APPEPPPPPPGVSGHINIPILILEELRVQQR 190
Db 392 ILQGLALENSGLSTIPAAAHPRNDSCHARTSAAPHACNPNDVSVSNEDOTDPTDAEQEK--E 449

```


191 QIHOQMTEOIC---ROVLLIGSLGQTVGAP-----ASPEL----- 224
 450 SPEKIDKKEKKMKIKKSPFLGSRRENGVAMHVCYCAKEKRPDVLVHRIHTEHK 509
 225 ---PQTGAASSTKPL---PLSPKPAQTGKTTASSSSSS-----SSGAEPKQA 270
 510 FKPCQCFRAFAVASTLAHKTHTGIRAKFCQYCKMSFSTGSLKVIRLHTGRP---- 565
 271 FEHLVPLGSGHPFSGVGVRSH-----KP-----PAPSPALPGS 306
 566 -PACPH---CDKKFRTSGHKKHTASHEKTELKRMHQRKPAKVNGKTNVPDPLPQ 621
 307 TDOLIASPHLAFPGTGLLAQCSLGAAGLEAASPGILPKNGSGELGY-GEVYISLEK 365
 622 EPLITDGLIQ-----IPKQFQSYNNNNVNNENDR 655
 366 PGRHKRCRCACAFVGSLSALQILHRSHTGERPYKCNVCSNRPFTTGNLKVHFRH---- 420
 656 P---YKCFYCHRAYKKSCHLKQHIRSHTEKPKFCSCQGRGFVSAGVLKANHIRTHTGLKS 712
 421 -----REKYPVQMNPVREHLDYITSSGLP-----YCMSPRE 456
 713 FKCLICNGAFITGGSLSRHHNGIHNDRMPCRYCQKTEKTSLNCKHMKTHRYELAQQL 772
 457 KAEERAGTPGGVERKPLVASTALASATESLTLSTGTAVAPGLPTENKFVLMKAVEP 516
 773 OHQQAASIDSTVYDOOSMASTOMOVEIESEDELPTOTAEVYAANP-----EAMLDLEP 824
 517 KS-----KADENTPPSGSGSALAGVADSGSATINQSLKVLTSLSMA 558
 825 QHVNGTEEAGLGOQLADPLEADEDEGFAVADPLRGHVDFEBSQPAQCSFEPAGLPQGF 884
 559 LLTNLHKSFGSPFPVYLEPLGASPSERSKIQQLYEKIDRQGVANAASTASGATTSAPA 618
 885 TVTDYVYHQQRPFPVYQQLQ-----DSTLESQALSTSFHQOSILOAPSSQMNVTTLRIQ 939
 619 PSS-----SASGPN-----QCVYCLRVLSGPRALRLHYGQHGGERPKCKV 659
 940 ESSQBELDIAQOSSQPLEDENEDRSRYRCQYCNKGFKKSHLKQHVRSHTGEKPYCKKL 999
 660 CGRAFSTRGNLRAHFVGHKTSPPARAONSCRICQKFTNAVTLQOHVRLHMLGQIPNGGS 719
 1000 CGRGFVSGVLKSH---EKHTGVKVA--FSCSYCNASFTTNGSLTRHATHTMSMK--PYKCP 1054
 720 ALSEGGCA---AQENSEQGTASGSPFPQSQSPREEMSEEEEDD---EEDEEDVTD 774
 1055 FCEGGRTTVHCKKHKHMRHQTV-----PSAVSANGETEGDTCOMEEEEEEESD 1101
 775 EDSLARGSESGEKAISVRGDSSEVSGAEDEVATSVAAPTTVKEMDSNEKAPQHTLPP 834
 1102 RNA-----SRKSRPEVITPTEBTA----- 1121
 835 PPPDNLDHPQMEQGTSD---VSGAMEEAKLIGISSPMALTOEGECTS--TPLEE 888
 1122 ---QIARIPQESATVSEKVLQVQAAEKD--RISELARDKQALODEPRKANCTCYCPKS 1175
 889 LNLPEMKKDKGSSSGK--ACEVCGOSFPQTALAEHOKTKHKDGLFQCVRQGLD 946
 1176 FKPSDVLVHVRIRHTGKRYKCDGCKSFYKSTLDCHVKTHTGQ--KLFSCHVCSNAST 1234
 947 RATLKKHMLAHNQVPPF-APV 967
 1235 KGSLLVHML-HTGAKRFKCPH 1255

GN ZNF341.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 NC NCB1:TaxID=9606;
 RX MEDLINE=21638749; PubMed=11780052;
 RA Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R.,
 RA Jones M., Stavrides G., Almeida J.P., Babbage A.K., Bagguley C.L.,
 RA Bailey J., Barlow K.F., Bates K.N., Beard L.M., Beare D.M.,
 RA Beasley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J.,
 RA Buck D., Burrill W.D., Butler A.P., Carder C., Carter N.P.,
 RA Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M.,
 RA Clegg S., Cobley V.E., Collier R.E., Connor R.E., Corby N.R.,
 RA Coulson A., Coville G.J., Deadman R., Dhami P.D., Dunn M.,
 RA Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,
 RA Graffham D.V., Griffiths C., Griffiths M.N.D., Gwilliam R., Hall R.E.,
 RA Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J.,
 RA Huckle E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D.,
 RA Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,
 RA Lehasalaio M.H., Leyerisha M.A., Lloyd C., Lloyd D.M., Lovell J.D.,
 RA Marsh V.L., Martin S.L., McConachie L.J., McElay K., McMuray A.A.,
 RA Milne S.A., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T.,
 RA Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I.,
 RA Phillimore B.J.C.T., Prathalingam S.R., Plumb R.W., Ramsay H.,
 RA Rice C.M., Ross M.T., Scott C.E., Sehra H.K., Shownkeen R., Sims S.,
 RA Stuce C.D., Smith M.L., Soderlund C., Steward C.A., Sulston J.E.,
 RA Swan R.M., Sycamore N., Taylor R., Tee L., Thomas D.W., Thorpe A.,
 RA Tracey A., Tromans A.C., Vaudin M., Wall M., Wallis J.M.,
 RA Whitehead S.L., Whitaker P., Willey D.L., Williams L., Williams S.A.,
 RA Wilming L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,
 RA Rogers J.;
 RT "The DNA sequence and comparative analysis of human chromosome 20.";
 RL Nature 414:865-871(2001).
 RN [2]
 RP SEQUENCE OF 227-773 FROM N.A.
 RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
 RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,
 RA Magatsuma M., Hosoi T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
 RA Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
 RA Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,
 RA Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K., Masuno Y.,
 RA Niimura Y., Iwayanagi T.;
 RT "NEDO human cDNA sequencing project.";
 RL Submitted (May-2001) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: MAY FUNCTION AS A TRANSCRIPTION FACTOR.
 CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).
 CC -1- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-FINGER PROTEINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AL050349; CAC14087.2; -;
 DR EMBL; AK027550; BAB5193.1; -;
 DR HSSP; P07248; 2ADR.
 DR Genew; HGNC:15992; ZNF341.
 DR InterPro; IPR000822; Znf_C2H2.
 DR Pfam; PF00096; Zf-C2H2; 13.
 DR SMART; SM00355; Znf_C2H2; 13.
 DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 9.
 DR PROSITE; PS0157; ZINC_FINGER_C2H2_2; 12.
 KW Transcription regulation; Zinc-finger; Metal-binding; Nuclear protein;
 KW DNA-binding; Repeat.
 FT ZN_FING 42 65 C2H2-TYPE (ATYPICAL).
 FT ZN_FING 227 251 C2H2-TYPE (ATYPICAL).
 FT ZN_FING 269 291 C2H2-TYPE.

816 TVKEM-DSNEKAQHHTLP PPPPPDLHDHPMEGQTSVSGAMEEAKL-----EGIS **868**

DR PRINTS; PR00048; ZINCfinger.
DR Prodom; PD000003; ZnF_C2H2; 7
DR SMART; SM00431; LER; 1.
DR SMART; SM00355; ZnF_C2H2; 13.

```
DR  SM00535; ZINC_C2H2; 13.  
DR  PROSITE; PS50804; SCAN_BOX; 1.  
DR  PROSITE; PS00028; ZINC_FINGER_C2H2_1; 13.
```


PROSITE: PS0157; ZINC_FINGER_C2H2_2; 13.
Transcription regulation; DNA-binding; Zinc-finger; Metal-binding;
Nuclear protein; Repeat; Alternative splicing.
DOMAIN 44 125
DOMAIN 310 321
DOMAIN 356 371
DOMAIN 356 378
DOMAIN 384 406
DOMAIN 412 434
DOMAIN 440 462
DOMAIN 463 484
DOMAIN 485 507
ZNFING 513 535
ZNFING 541 563
ZNFING 569 591
ZNFING 597 619
ZNFING 625 647
ZNFING 653 675
ZNFING 681 703
ZNFING 709 731
ZNFING 731 749
VARSPIC 250 257
CONFLICT 304 305
SEQUENCE 734 AA; 82036 MM; 2BE7D69B18F29437 CRC64;
AL -> RV (IN REF. 1).
EAGGIESP -> MNGPLVYA (IN ISOFORM MZFI/C).
Query Match 5.9%; Score 311; DB 1; Length 734;
Best Local Similarity 19.4%; Pred. No. 5.9e-07;
Matches 204; Conservative 74; Mismatches 330; Indels 444; Gaps 39;
11 LGPGCEPERERG-----DASEEH-----POVCA--KC--CAQSPDTEFLAHQ 51
6 LGSPRAPPEDECPVAVKLEDESEGEALMDPGPEARLRRCFYEATGQALAQ 65
52 NSCCTDPVAVIIGOEPPNSASSAPRPEGHRSQVMDT-----EHSNPD--SGSS 103
66 RELC-----RQMLRPEVRSKQEMLELVLEQFLGALPEIQARVQ 105
104 GPPDPWGPERRGEESGQFLVATGTAAGG-----GGLIASPKIGAT--PLPP 152
106 G-----QRPSPPEAALVLDGLRRPGPRRWVTVVOGQEVLEKPESSFOPLPE 157
153 ESTAPPPPPPPPPPGVSGHLNPLLEELRVLOQROI-----HOMOMTEQICROVL 207
158 TEPPEPPPPKPP-----RTQESPLGLQVKESEVTD----- 192
208 LGLSQTVGAPASPSBELGTGAASSTKPLDLPSPIKPAQGTKTASSSSSSSGAEP 267
193 -----SDPLESGPLAATQESVPLLP-EAQRGCTVLDIQIPHSKTPGEGP 237
268 KQAFHLYPLGSHPPSVGVGSHKTPAPSPAL-----PGSTDQLINPPLAFP 319
238 SWR-----EHPRALMHE-ENAGGIF-----SPGFALQIGSISAGPS-----VSPHLHP 280
320 GTTGL--LAAQCLGAARLEAAASPGLLKPKNGSGE-----LGYEIVSISLEKP 366
281 WDLGAGLGSQIOSPSR--EGGFHALLPLDLRSQDPDTDEDPCRGVALLITRMWSP 338
367 GGRH-----KCRFCAKVFSGDSALQIHLRSHGTGPRPKCANNCGNFTTNGN 412
339 RGRSRGRPTGGGVNRGRCRDYCGVFQSRNMLRHQKITHTGERPFVCSDEGRSFRSRS 398
413 LKVHPIRHRKEVHYVQNNPHVPEHLIDVITSSGLPYGMSVPERKEAEAGTPGGVBRK 472
399 LIRHQLTHTEERFV-----GDDCGG----- 420
473 PLVASTTALSATSTLTLSTGTAVADGLPTFNKFLMKAVEPKSKADENTPPGSEGA 532
421 -----FVRSARLEHRVHNGEOP----- 439
533 IAGVADSGSATRMQLSKLVWISPLSWALLTNHLKSTGSPFPYVLEPIGASPSFTSKLOL 592
440 -FRACBGGGFRGRSN-----LLOHORIHGDPPGP-----GAKP----- 472

QY 593 VEKIDRGAVAVASTASGAPTSAPAPSSA-----SGPNO--CVICL 633
Db 473 -----PAPGAPDEPPGPPPCSECRSEFARRAVILEQAVITGDKSPGCEVG 519
QY 634 RVLSCPRALRLHYGQNGGERPRPKVCGRASTRGNLRAHFVGHKTSPPAARQNSCPICQ 693
Db 520 ERGRRSRSLTLQHRVHSGRFPAPCAECGGSFQRSNLTQHRRIH-----TGERPFACAE 575
QY 694 KKFTNAVTLQOHVHRNHLGQIIRNGSALSSEGAQDSSQSTASGSPFPQSQOPS 753
Db 576 KAFQRPRLTQHLRVHTG-----EKP- 596
QY 754 PEELMSEEEDEEEDVTDDESLAGSGSEGEKALSVGRDSEEVSGAEEVATSVAA 813
Db 597 -----FAC 599
QY 814 PTVKEMDSNEKAPQHTLPPPPPPNDLHPQMEQSTSDVSGAMEEAKLEGISSPMAA 873
Db 600 PEGGGRFSGRLKTRHQRHTGKRP--YHCGEGGLGFTQVS----- 638
QY 874 LTQGEGETSTPLVEELNLPKAKKDPGESSGRKACEVCGGSEFPOTALEHOKTHPKDP 933
Db 639 -----RLTEHORIHTEGER--FACPGCGSFRQHANLTQHRRIHTEGERP 680
QY 934 LTGVCFCRQGFIDRATLKKHMLAHHQVPPA 965
Db 681 -YACPECGAKAFQRPRLTQH-LFTHREKPEFA 710
RESULT 12
2133_HUMAN
ID 2133_HUMAN STANDARD; PRT; 654 AA.
AC P52736; O9H443; O9BUV2;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
GN Zinc finger protein 133.
ZNF133.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Insulinoma;
RX MEDLINE=95377390; PubMed=7649249;
RA Vissing H., Meyer W.-K., Aagaard L., Tommerup N., Thiesen H.-J.;
RT "Repression of transcriptional activity by heterologous KRAA domains
present in zinc finger proteins.";
RL FEBS Lett. 369:153-157(1995).
[2]
RP SEQUENCE FROM N.A.
RX MEDLINE=21638749; PubMed=11780052;
RA Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R.,
RA Jones M., Stavrides G., Almeida J.P., Babbage A.K., Bagguley C.L.,
RA Bailey J., Barlow K.F., Bates K.N., Beard L.M., Beare D.M.,
RA Beasley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J.,
RA Buck D., Burrill W.D., Butler A.P., Carder C., Carter N.P.,
RA Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M.,
RA Clegg S., Cobley V.E., Collier R.E., Connor R.E., Corry N.R.,
RA Coulson A., Coville G.J., Deadman R., Dhami P.D., Dunn M.,
RA Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,
RA Grafham D.V., Griffiths C., Griffiths M.N.D., Gwilliam R., Hall R.E.,
RA Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J.,
RA Huckle E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D.,
RA Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,
RA Levaslainho M.H., Leverisa M.A., Lloyd C., Lloyd D.M., Lovell J.D.,
RA Marsh V.L., Martin S.L., McConachie L.J., McEay K., McMurray A.A.,
RA Mine S.A., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T.,
RA Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I.,
RA Phillimore B.J.C.T., Prathalingam S.R., Plumb R.W., Ramsay H.,
RA Rice C.M., Ross M.T., Scott C.E., Sefira H.K., Showkeen R., Sims S.,

[illegible]

Db	225	APPPQAPQPPRPPEPERSVQPEPEPEEREA-----IPTAPAPADEEPAPPE--	273
QY	630	VICLRVLCSPRALRLHYGHGGERPFCKVCYCGRAFSTGNLRANHFVGHNTSPARAQNSC	689
Db	274	-----FRQYVCGQSPFQSWFLTGHHMKHKAS----FDHAC	304
QY	690	PLCCQKFTANVTQIQVHRMLGGQIPNGSALSBEGGAAQEMSSDQSTASPGSPFPQ-	748
Db	305	PVCGRCFEKPEWFLKNHMKVHASKIGP-----LRAPG-----PASGPAPAPQPD	348
QY	749	-----SQCP-----SPEEEMSEEEED-----EEEEEDVTDSDLAGRSESGG--	787
Db	349	LGILATVEPLGALLIAPAPTPAPERRPPSLCYGLSLRAGEGRPNGBAERPPRSFGFR	408
QY	788	--EKAISVNG-----DSEEVSGAEEEVATSVAAPPTVYKEMDSNEKAPQHTLPPPP	836
Db	409	PLSSALPARARRHRAPEEPEEEVEVAQEEFTWARGSLGSLAST-----	452
QY	837	PRPDNDHPQMEGTSIVSGAMEEAEKLEGISSPAALTOEBEGSTPLVEELNLPBAMK	896
Db	453	-----HPRGFE-GPGHSASAGAQAQ-----STATQEBNGL-----LVGGR	488
QY	897	KDPGESSGRACACVCGSPFTQPTALEBHOKTNPKDPLETCVFCRQGFDRATIKKMLL	956
Db	489	PEGGRGATGKDCFPCKGSPFSANHLKHLRHTGRRP-YKCPHCDYAGQSSGLKHN-LQ	546
QY	957	AHH-----QVPPFAPRHG-PQNTATLSLP-----GCSS-SIPFGLSPFPR	995
Db	547	RHHREORGAGPPEPPPPRGSGAPQSGAKPSPQATWEGASPPRPSSGACPGSR	606
QY	996	KDDPTMP 1002	
Db	607	R-KPASP 612	
RESULT 14			
ID	_Y296_HUMAN	STANDARD;	PRT: 1829 AA.
AC	015015;		
DT	16-OCT-2001 (Rel. 40, Created)		
DT	16-OCT-2001 (Rel. 40, Last sequence update)		
DT	15-JUN-2002 (Rel. 41, Last annotation update)		
DE	Hypothetical zinc finger protein KIA0296.		
GN	KIA0296.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.		
OX	NCBI_TaxID=9606;		
RP	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=Brain;		
RX	MEDLINE=97349984; PubMed=9205841;		
RA	Nadase T., Ishikawa K.-I., Nakajima D., Ohira M., Seki N.,		
RA	Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;		
RT	"Prediction of the coding sequences of unidentified human genes. VII.		
RT	The complete sequences of 100 new cDNA clones from brain which can		
RL	code for large proteins in vitro.;"		
RL	DNA Res. 4:141-150(1997).		
CC	-1- FUNCTION: MAY FUNCTION AS A TRANSCRIPTION FACTOR.		
CC	-1- SUBCELLULAR LOCATION: Nuclear (Probable).		
CC	-1- SIMILARITY: BELONGS TO THE KRUPEL FAMILY OF C2H2-TYPE ZINC-		
CC	FINGER PROTEINS.		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -		
CC	the European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/		
CC	or send an email to license@isb-sib.ch).		
CC	-----		
CC	EMBL: AB002294; BAA20756.1; -.		

	DR	InterPro: IPR000822; Znf_C2H2.
	DR	Pfam: PF00096; Zf-C2H2; 31.
	DR	PRINTS: PRO0048; ZINC_FINGER.
	DR	SMART: SM00355; Znf_C2H2; 30.
	DR	PROSITE: PS00028; ZINC_FINGER_C2H2_1; 30.
	DR	PROSITE: PS00157; ZINC_FINGER_C2H2_2; 29.
	KM	Hypothetical protein: Transcription regulation: DNA-binding;
	KW	Zinc-finger; Metal-binding; Nuclear protein; Repeat.
FT	DOMAIN	8 1783 ZINC FINGERS.
FT	ZN_FING	8 31 C2H2-TYPE.
FT	ZN_FING	48 70 C2H2-TYPE.
FT	ZN_FING	75 97 C2H2-TYPE.
FT	ZN_FING	239 261 C2H2-TYPE.
FT	ZN_FING	266 288 C2H2-TYPE.
FT	ZN_FING	294 316 C2H2-TYPE.
FT	ZN_FING	374 396 C2H2-TYPE.
FT	ZN_FING	401 424 C2H2-TYPE.
FT	ZN_FING	465 487 C2H2-TYPE.
FT	ZN_FING	492 514 C2H2-TYPE.
FT	ZN_FING	575 597 C2H2-TYPE.
FT	ZN_FING	617 639 C2H2-TYPE.
FT	ZN_FING	644 666 C2H2-TYPE.
FT	ZN_FING	821 843 C2H2-TYPE.
FT	ZN_FING	848 870 C2H2-TYPE.
FT	ZN_FING	881 904 C2H2-TYPE.
FT	ZN_FING	958 980 C2H2-TYPE.
FT	ZN_FING	1052 1074 C2H2-TYPE.
FT	ZN_FING	1079 1101 C2H2-TYPE.
FT	ZN_FING	1203 1225 C2H2-TYPE.
FT	ZN_FING	1230 1252 C2H2-TYPE.
FT	ZN_FING	1258 1280 C2H2-TYPE.
FT	ZN_FING	1299 1321 C2H2-TYPE.
FT	ZN_FING	1364 1386 C2H2-TYPE.
FT	ZN_FING	1557 1579 C2H2-TYPE.
FT	ZN_FING	1585 1607 C2H2-TYPE.
FT	ZN_FING	1677 1699 C2H2-TYPE.
FT	ZN_FING	1704 1726 C2H2-TYPE.
FT	ZN_FING	1732 1754 C2H2-TYPE.
FT	ZN_FING	1761 1783 C2H2-TYPE.
SQ	SEQUENCE	1829 AA; 200760 MW; B2OC48B38684895 CRC64;
	Query Match	5.78; Score 301; DB 1; Length 1829;
	Best Local Similarity	20.0%; Pred. No. 3.9e-06;
	Matches 168; Conservative	87; Mismatches 291; Indels 296; Gaps 34;
OY	244 KPAONGKTTAASSSSSSSSCAEPPKQAFFHLXHPGLGSOHPFSGVGYGRSHKPTTPAPSPAL	303
Dd	796 KPA--TGOPNPSHHSAANVTGMQ-----AGAAHTCSDCG-----	827
OY	304 PGSTDLIASPHLAFPGTTCGLLAO--CLGAARGLEMAAASPGLLKPKNGSGELGEVISS	362
Dd	828 -----H-SFRPATGLLSHRPC-----	842
OY	363 LEKPGGRHKCRFCAKAYGVGSALQIHLRSHTGERPYKCNCVGNNPRTTRGNLVHF	417
Dd	843 --HPRGIYCCLCPKEEDSLPALRSFHQNHRPGFAATSAQPLCCLCGMIRPRAGYRLH-	899
OY	418 HRRREKYRVHQMPRHVRPEHLDIYTSSGLRGYMSVPPEKAEEEAETPGGCVGEKKPLVAS	477
Dd	900 ---RROAH-----SSSGMEG-----SEEEEGEVAEAAPARSP	931
OY	478 TTALSATVESTLLSTGTAVAPGLPTRENKFVLIMKAVEPKSKADEMTPRSEGSALINGVA	537
Dd	932 PLQLSASELLN-----QLQREVELADSDAGYGHCCCGOT---YD	968
OY	538 DSGSATRMOLSKLVTLSPSWALLTNHLKSTGSEFPPEYLVEPLGASEP-----TSKIQL	592
Dd	969 DLGSLER-----HHQSOSGDTADKAPSPLGVAGDAMEVMVDVSLEDI	1011
OY	593 VEKIDOGAANAATASTACGAPTTPASAPASSASGAPNOVCYLCRLVSCPRAIRLHGONIGE	652
Dd	1012 VNSVSDEGGDAKQEGEGTLP-----GDISLCTI-----OGGE	1042

Oy	653	- - - - -	-RPFKCVGSRASTSTKCNLNAHPGHNTSPAAKQNSCPICQKKTNAVTLDOH	705	
Db	1043	SLLLEAPREFRCNOCKRTYRHGSLVNHRRKIQTGFEL- - - - -	CPVSRCCYPNLAAAYRNH	1097	
Oy	706	VAMH- - - - -	LGGQIPN--GGALSSEG- - - - -	GGAQENSSSEOSTASGPCSFPPOSOQ	751
Db	1098	LNHHPRCKSSSEQVGIFPPEAGSSSELQVGFIPREGS- - - - -	KRPQHMAEEGRG- - - - -	QAIV	114.9
Oy	752	PSPEEMSEBE- EDEDEEEDVTDEDLSLAGSGE- - - - -	SGGEKAISVRGDSE- - - - -	EVSIG	802
Db	1150	EXLOEBLKVEPLIEEVARVAKEEVEETTVKGEIIEPLRETAKEKGCGTAEASSRPPSCVCG	12099		
Oy	803	AEEFATSV- - - - -	-AAPTYVKENDSNKAPQHLLPPPPPDDNIIDHQPMEGS	850	
Db	1210	RSYKHAGSLINHHROSHQTGHFGCQACSKGSNNMLSKNH- - - - -	RRIHADPRFR	12595	
Oy	851	TSDVSGAMEEAKLECISPPMALTOGBESTPIRYE- - - - -	LNLPPEARK	897	
Db	1260	CSECGKAFILRKQL- ASHQRVIMERRGGCGTKATREDPRFGCGCRTRYAHAGSLINH	13177		
Oy	898	DGESSGRACEVCGSFPTOTALLEBHKHTPKD- - - - -	GPLTFVCFCRGFLDRA	948	
Db	1318	RSHETGYQSCTPCFTYSNBALKDHQRHLSENRRBRAGRSRTAVRCAALCGHSFPGG	13778		
Oy	949	TLKKHMLLNHHOVPAPP- - - - -	-HEPONIAI- - - - -	LSLVPGCSSSIPTS	987
Db	1378	SLERR- LREHETEREPANOGGLDGTAAASEANLTGSGILETQLGGAEPVPHLEDGVR	14355		
Oy	988	PG 989			
Db	1436	PG 1437			
 RESULT 15 ZN84_HUMAN STANDARD: PRT: 738 AA.					
ID	ZN84_HUMAN				
AC	P51523; Q9NNX7; Q9UC17; Q9UC18;				
DT	01-OCT-1996 (Rel. 34, Created)				
DT	01-OCT-1996 (Rel. 34, Last sequence update)				
DT	15-JUN-2002 (Rel. 41, Last annotation update)				
DE	Zinc finger protein 84 (Zinc finger protein HPF2).				
GN	ZNF84.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.				
OX	NCBI_TaxId=9606;				
NP	[1]				
RP	SEQUENCE FROM N.A.				
RC	TISSUE=Placenta;				
RX	MEDLINE=89377476; PubMed=2505992;				
RA	Bellefroid E.J., Decocq P.J., Benhida A., Poncellet D.A.,				
RA	Belayev A., Marital J.A.;				
RT	"The human genome contains hundreds of genes coding for finger				
RL	proteins of the Kruppel type.";				
RL	DNA 8:377-387(1989).				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RA	Grimaldi G.;				
RL	Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.				
RN	[3]				
RP	SEQUENCE OF 1-257 FROM N.A.				
RC	TISSUE=Teratocarcinoma;				
RX	MEDLINE=92051312; PubMed=1945843;				
RA	Rosetti M., Marino M., Franze A., Tramontano A., Grimaldi G.;				
RT	"Members of the zinc finger protein gene family sharing a conserved N-				
RL	terminal module.";				
RL	Nucleic Acids Res. 19:5661-5667(1991).				
CC	-I- FUNCTION: MAY BE INVOLVED IN TRANSCRIPTIONAL REGULATION.				
CC	-I- SIMILARITY: BELONGS TO THE KRUPPEL FAMILY OF C2H2-TYPE ZINC-				
CC	FINGER PROTEINS.				
CC	-I- SIMILARITY: CONTAINS 1 KRAA DOMAIN.				

EMBO J. 6:3065-3070(1987).
[2]
CHARACTERIZATION.
MEDLINE-94021366; PubMed-7692399;
Andrezzioli M., de Lucchini S., Costa M., Barsacchi G.;
"RNA binding properties and evolutionary conservation of the Xenopus
multifinger protein Xfin.";
Nucleic Acids Res. 21:4218-4225(1993).
[3]
STRUCTURE BY NMR OF FINGER 31.
MEDLINE-89346749; PubMed-2503871;
Lee M.S., Gipeert G.P., Soman K.V., Case D.A., Wright P.E.;
"Three-dimensional solution structure of a single zinc finger DNA-
binding domain.";
Science 245:635-637(1989).
[4]
STRUCTURE BY NMR OF A FINGER.
MEDLINE-89378224; PubMed-2506074;
Lee M.S., Cavanagh J., Wright P.E.;
"Complete assignment of the 1H NMR spectrum of a synthetic zinc
finger from Xfin. Sequential resonance assignments and secondary
structure.";
FEBS Lett. 254:159-164(1989).
-1- FUNCTION: BINDS RNA. COULD FUNCTION IN POST-TRANSLATIONAL
REGULATION PROCESSES.
-1- SUBCELLULAR LOCATION: CYTOPLASMIC.
-1- TISSUE SPECIFICITY: OOCYTES AND IN SPECIALIZED CELL TYPES SUCH
AS NEURAL RETINA CONES.
-1- DEVELOPMENTAL STAGE: EXPRESSED THROUGHOUT OOGENESIS AND
EMBRYOGENESIS.
-1- DOMAIN: CONTAINS 37 FINGER MOTIFS IN 6 DOMAINS.
-1- PTM: PHOSPHORYLATED.
-1- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-
FINGER PROTEINS.
-1- SIMILARITY: CONTAINS 1 KRAB DOMAIN.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL: X06021; CAA29425.1; .
PIR: S00647; S00647.
PDB: 1ZNF; 15-OCT-91.
InterPro: IPR001909; KRAB.
InterPro: IPR000822; ZnF_C2H2.
Pfam: PF01352; KRAB; 1.
PRINTS: PR00048; ZINCINGER.
ProDom: PD000003; ZnF_C2H2; 20.
SMART: SM00349; KRAB; 1.
SMART: SM00355; ZnF_C2H2; 35.
PROSITE: PS00805; KRAB; 1.
PROSITE: PS00028; ZINC_FINGER_C2H2_1; 35.
PROSITE: PS0157; ZINC_FINGER_C2H2_2; 37.
Zinc-finger: Metal-binding; RNA-binding; Repeat; 3D-structure;
Phosphorylation.
DOMAIN 1 58 KRAB.
DOMAIN 108 298 ZINC-FINGERS I.
DOMAIN 326 488 ZINC-FINGERS II.
DOMAIN 503 721 ZINC-FINGERS III.
DOMAIN 750 940 ZINC-FINGERS IV.
DOMAIN 988 1066 ZINC-FINGERS V.
DOMAIN 1136 1298 ZINC-FINGERS VI.
ZNF_FING 108 130 C2H2-TYPE.
ZNF_FING 136 158 C2H2-TYPE.
ZNF_FING 164 186 C2H2-TYPE.
ZNF_FING 192 214 C2H2-TYPE.
ZNF_FING 220 242 C2H2-TYPE.
ZNF_FING 248 270 C2H2-TYPE.

FT	ZN_FING	276	298	C2H2-TYPE.
FT	ZN_FING	326	348	C2H2-TYPE.
FT	ZN_FING	354	376	C2H2-TYPE.
FT	ZN_FING	382	404	C2H2-TYPE.
FT	ZN_FING	410	432	C2H2-TYPE.
FT	ZN_FING	438	460	C2H2-TYPE.
FT	ZN_FING	466	488	C2H2-TYPE.
FT	ZN_FING	503	525	C2H2-TYPE.
FT	ZN_FING	531	553	C2H2-TYPE.
FT	ZN_FING	559	581	C2H2-TYPE.
FT	ZN_FING	587	609	C2H2-TYPE.
FT	ZN_FING	615	637	C2H2-TYPE.
FT	ZN_FING	643	665	C2H2-TYPE.
FT	ZN_FING	671	693	C2H2-TYPE.
FT	ZN_FING	699	721	C2H2-TYPE.
FT	ZN_FING	750	772	C2H2-TYPE.
FT	ZN_FING	778	800	C2H2-TYPE.
FT	ZN_FING	806	828	C2H2-TYPE.
FT	ZN_FING	834	856	C2H2-TYPE.
FT	ZN_FING	862	884	C2H2-TYPE.
FT	ZN_FING	890	912	C2H2-TYPE.
FT	ZN_FING	918	940	C2H2-TYPE.
FT	ZN_FING	988	1010	C2H2-TYPE.
FT	ZN_FING	1016	1038	C2H2-TYPE.
FT	ZN_FING	1044	1066	C2H2-TYPE.
FT	ZN_FING	1136	1158	C2H2-TYPE.
FT	ZN_FING	1164	1186	C2H2-TYPE.
FT	ZN_FING	1192	1214	C2H2-TYPE.
FT	ZN_FING	1220	1242	C2H2-TYPE.
FT	ZN_FING	1248	1270	C2H2-TYPE.
FT	ZN_FING	1276	1298	C2H2-TYPE.
FT	STRAND	1045	1045	
FT	STRAND	1052	1052	
FT	HELIX	1056	1065	
FT	TURN	1066	1066	
SEQUENCE	1350 AA; 155804 MW; 27F10AB0851E0AD8 CRC64;			
Query Match	5.5%; Score 290; DB 1; Length 1350;			
Best Local Similarity	20.9%; Pred. No. 8.6e-06;			
Matches 126; Conservative 61; Mismatches 248; Indels 168; Gaps 18;				
QY	364	EKPGRHRCFRCAKVGSDSALQHLMSHGERPKYKCVGCGRRFTTRGKLVKVFHRRREK	423	
DB	407	EKP---FKCSHCDKRTKTERALKHQRTHGEKPKSCDCKEFTORSNLIHQRIHTE	463	
QY	424	YPHVQMNPRVPEHLDVITSSGLPYGMSVPEKAEBAAGTGGVERRDLVASTALSA	483	
DB	464	RPY-----	466	
QY	484	TESTLLSTGTSTAVAPGLPTFNKFLMKAVEPKSKADENTPPGSEGSALAGVADSGSA	542	
DB	467	--KCTLCD-----RTFIQNSDLVKHQKVHANLP-----LSDEPTAN	500	
QY	543	TRMQLSKIVTSLPSMALLTNHLK-STGSEFPFVYLEPAGASPESTKLQVLVEKIDRGA	601	
DB	501	SPHKSCDLTFPSHMTFKMKSHLHSEKRP-----QCAECK--GFTOKSDLVKH	549	
QY	602	VAVASTAGAPTTASAPSSASGPNOCVICTLRVLSCPRALRLHYGOHSGSERPKCKVCG	661	
DB	550	IRV-----HTGEKPFCKILCKKSSQNSDLKHMRIHNGEKFPFCYTCD	593	
QY	662	RAFSTRGNIRAHFVGKTSFPAARQNSCPICQCKFTNAVTLQOHVHMLGGQILPNGSAL	721	
DB	594	KSFTERSALIKH--HRTHTGER-PHKCSVCQKGFIOKSLTKHSRTHNG-----	639	
QY	722	SEGGAAQENSSQSTASGSGSPFOPOSQSPSEPFMEDEEEDDEEDVDNEDSLAGR	781	
DB	640	-----EKPYPCQCGK-SFQNSDLVKHQRTHGEKPNHTECNKRFTEGSSLVKH	689	
QY	782	GSESGEKALISVRGDEEVSAGAEBAEVATVAPTTVKEMDSNEKARQHTLPPPPPP	838	
DB	690	RTHSGEK-----PYRCPOCEKFTIQSSDLVKHLVHNGENPPATAFH	733	

837 KCKSLSEIOISARANGNNSPMNSDAMEKDESDADAPHSDDTTSVPSPPLHSSSIVA 896
856 GAMESEAKLEGTSSPAALAQEGEGSTSPRIEELNLEPMKRPSSSGKACVCGQSF 915
897 PIPITPPQNEFLQSIILAQASLIG-----PLT--ANRPSAF-----YCHCKIPF 938
916 PQTALAEHQKTHPKDGLFTCVCFROGFLDRATLKKHMLLAHQ 960
939 DTQVLDLSHMRFTPCNP-FMCSDCQYQAFNELSPALHMYQARHQ 982
RESULT 19
EPI_MOUSE
D ZEP1_MOUSE STANDARD; PRT; 2688 AA.
C 003172:
T 01-OCT-1996 (Rel. 34, Created)
T 01-OCT-1996 (Rel. 34, Last sequence update)
T 15-JUN-2002 (Rel. 41, Last annotation update)
T Zinc finger protein 40 (Transcription factor alphanA-CRYBP1) (Alpha A-crystallin-binding protein 1) (Alpha A-CRYBP1).
T HIVEP1 OR ZNF40 OR CRYABP1.
T Mus musculus (Mouse).
T Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
T NCBI_TaxID=10090;
T [1]
P SEQUENCE FROM N.A.
P SPRAIN-DBA/21:
C MEDLINE=95138112; PubMed=7836383; Donovan D.M., Platiorsky J., Brady J.P., Kantorow M., Sax C.M., "Murine transcription factor alpha A-crystallin binding protein I. Complete sequence, gene structure, expression, and functional inhibition via antisense RNA." J Biol. Chem. 270:1221-1229(1995).
T [2]
P SEQUENCE OF 2024-2688 FROM N.A.
P TISSUE=Lens epithelium;
C MEDLINE=90287161; PubMed=1694016; Nakamura T., Donovan D.M., Hamada K., Sax C.M., Norman B., Flanagan J.R., Ozato K., Westphal H., Platiorsky J., "Regulation of the mouse alpha A-crystallin gene: Isolation of a cDNA encoding a protein that binds to a cis sequence motif shared with the major histocompatibility complex class I gene and other genes." Mol. Cell. Biol. 10:3700-3708(1990).
T -1- FUNCTION: TRANSCRIPTION FACTOR WHICH BINDS SPECIFICALLY TO THE PALINROMIC SEQUENCE 5'-GGGAATGCC-3' IN THE ALPHA-A CRYSTALLIN PROMOTER.
T -1- SUBCELLULAR LOCATION: Nuclear.
T -1- DOMAIN: CONTAINS TWO SETS OF 2 ZINC-FINGERS, WHICH ARE WIDELY SEPARATED AND RECOGNIZE THE SAME DNA SEQUENCE. THERE IS A FIFTH ZINC-FINGER IN-BETWEEN.
T -1- SIMILARITY: STRONG TO HIVEP2.
C This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
C
C EMBL; L36829; AAA98810.1; JOINED.
C EMBL; L36825; AAA98810.1; JOINED.
C EMBL; L36836; AAA98810.1; JOINED.
C EMBL; L36827; AAA98810.1; JOINED.
C EMBL; L36828; AAA98810.1; JOINED.
C EMBL; X68946; CAA48762.1; JOINED.
C HSSP; P15822; 3ZNF.
C TRANSFAC; T00007; HIVEP1.
C MGD; MGI:96100; HIVEP1.
C InterPro; IPR000822; Znf_C2H2.
C Pfam; PF00096; zf-C2H2; 5.
C PRINTS; PR00048; ZINCFINGER.

DR SMART; SM00355; ZNF_C2H2; 4.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 4.
DR PROSITE; PS0157; ZINC_FINGER_C2H2_2; 4.
KW Transcription regulation; zinc-finger; Metal-binding; DNA-binding;
KW Nuclear protein; Repeat.
FT DOMAIN 165 170 POLY-SER.
FT DOMAIN 407 459 ZINC_FINGERS.
FT ZN_FING 407 429 C2H2-TYPE.
FT ZN_FING 435 459 C2H2-TYPE.
FT ZN_FING 799 802 POLY-SER.
FT DOMAIN 851 854 POLY-PRO.
FT ZN_FING 953 976 C2HC-TYPE (POTENTIAL).
FT DOMAIN 1482 1486 POLY-SER.
FT DOMAIN 2074 2126 ZINC_FINGERS.
FT ZN_FING 2074 2096 C2H2-TYPE.
FT ZN_FING 2102 2126 C2H2-TYPE.
FT DOMAIN 2179 2182 POLY-ASP.
SQ SEQUENCE 2688 AA; 288341 MW; 5EAD46C3A700BBE6 CRC64;
Query Match 5.3%; Score 283.5; DA 1; Length 2688;
Best local similarity 20.1%; Pred. No. 3.2e-05;
Matches 224; Conservative 130; Mismatches 409; Indels 353; Gaps 48;
QY 3 QETGSSSLGPGCGEPAERG-----DASEHHPOVCAKCAQFSDPTFFLAHONSCTD 57
DB 107 QFTKONGETPTGMTAESSESGDLVSPKRTSSPHORSELRMRRESESDPTR----- 155
QY 58 PPVMTIIGQENPNSNSASAPREGHRSQVMDTSHENPPSGSSGPDPTWGPERGE 117
DB 156 -----LSGLDQORSSSS-----SKARTDSECSPP--CSTTPSYT----- 192
QY 118 ESSGQFLVA-----TCTAAGGGGGLLASPKL-----GATPLPEESTPAP----- 158
DB 193 STAFVLLKAMEPELSTLSQKSSCAIKTEKLRPKTKVRSPEKLKNSSLDAPNATSPDLV 252
QY 159 ---PPPPPPPPPPGVSGHLNPLILEELRYLQORQIHOMQTEQICROVLLIGSQIV 215
DB 253 VESPCPCPTSYPVHVAVSTOKSEQVAAOCVSHLYSSODHLVPKLSQONQO--LPGHLGFT- 309
QY 216 GAPAPSELPFGGAASSTKRLPLFSPIKPAQOTGKTASSSSSSSSSGAEPKQAFPHLY 275
DB 310 -----GSTLNLTLESTK--LEPLYNIAVTVGLT-----SPRTQVTPPHQ----- 351
QY 276 HPLGSQHPFVSGVGGRSHKPTPAPSPALPGSTDLIASPHLAFPTTGLLAACLGARG 335
DB 352 -QMDVSPPLSVSPASSTQSP--PGR-----IYSAHVAVS-----YSQS 387
QY 336 LEAASPGLLKPRKNSGELGYEYISLEKPGGRHKRCAPKAVVFGSDALQHLRSHTGE 395
DB 388 VEQMSLLLRDQK-----PKKQKGYICEVCNCAKAPSVLLKHIRSHTGE 432
QY 396 RPKYCNVCGNRFRTREGNLKVFHFRREKYPH-YQNPMPHPLDVLVITSSGLPYGSVP 454
DB 433 RPYCVTCGFSFRTKSNL---YKHKSHAHITIKGLVLOPE-AGGLFLSQEPKALSVH 487
QY 455 ---PEKAEAGTPGGGVERKPL-----VASTALSAATESLTLSTGTSTAVAPGLPTF 505
DB 488 SDIEDSGEDESGLAQRQNNPCVKDLQPVQTMKVSPDESPLKILPSNDHVVGRF--- 544
QY 506 NKFVLMKAVEPKSKADENTPPGSEGSALAGVADSSAIRMQLSKLVTSIPSMALLTNHLK 565
DB 545 -----SSQDRPSPSOQAPTELPK----- 561
QY 566 STGSPFPVYLEPLGASPSFETSKLOQLVEKIDROGAVAVASVAGAPTSSAPASSASG 625
DB 562 -----VVHPVSNAPPLKTDCLQ-----VAN-----PNEELPSPQS----- 591
QY 626 PNOCVTCLRVLSCPRALRL-----HYGOHG--ERPFKCKVCGRAVSTGRNLRAHFVGHKT 679
DB 592 -----PRLLHVASILSHSASVSLMEDSCHOKDVIQSEKPDH----- 632
QY 680 SPAAQNSCPTQCKFTTAQVNTLQOHVBMILGGQIPNGSGALSSEGGGAQENSSEOSTAS 739


```

DR Db      755 SSLRKHKRIHTEKPPKCKECCGKAFFINSSTLTNRKHRIHGEE--KPKCECGCAKFFSNT 812
QY        589 LQQLEVKEDRGGAVAAVASTASGAFTTAPAPSSSASGPNOVCYLRLVSCPRALRLHYGQ 648
Db      813 LTK--HKTIHTG-----EKPYKCKECCGAKFKHSSALAKHKII 847
QY        649 HGGERRPCKCAGGRASTRGNLRAHFVGHKTSPARAQNOSCPICKKFTNAVTLIQDHYRM 708
Db      848 HAGEELLYKCEEGCAKATNOSNLTTKKIITHEKEKPSKEE---CDRAFIVSSTLTENHRI 903
QY        709 HLGGQIPNGSALSSEGGAAQENSSEOSTASGP-----GSFPPOSOOPSPREEEMSEE 761
Db      904 H-----TREKPKYCEEGCAKAFSQSHLLTHKRMHNGEK 936
QY        762 EEDDEDEEDVDVDEDLAGSGESGGEKAISVRGDSEEVSGAEEVAATSYAAPTYVKEMD 821
Db      937 PYKCEECGGAFAFSQSSTLTTHKIITHTGKRPYC-----EECGAKARFSSSTLTNE- 983
QY        822 SNEKAPQHITLPPPPPNDNHQPMBQGTSDVGSAEMEAEKLLEGISSPMALTLQEBEGT 881
Db      984 -----HKIHI-----TEKPKYCEEGCAFAFSQSSTLTTRHTRMHTGE 1019
QY        882 STPLVEE---LNLPAMKKDPPESSGRK--ACEYCGSGFPQTALLENOKTPKDDPLF 935
Db      1020 KPYKCEEGCAKAFNRSSKLTTHKIITHTGKRPYKCECGCAKFISSSTLNGKHRIHREKP-Y 1078
QY        936 TCVFROGFELDRATLKHHMLLAHQVP 962
Db      1079 KCCECGKAFPOSSTLTRHKRLHTEGKP 1105

RESULT 22
Z1169_HUMAN
ID          2169   STANDARD:         PRT;       512 AA.
CD          014929;
DEFINITION  15-Dec-1998 (Rel. 37, Created)
            15-Dec-1998 (Rel. 37, Last sequence update)
            16-Oct-2001 (Rel. 40, Last annotation update)
ACCESSION   ZNF169.
FEATURES    znc_finger protein 169 (Fragment).
ORIGIN      Homo sapiens (Human).
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
            NCBI_TaxId=9606;
            [1]
SEQUENCE FROM N.A.
MEDLINE=97225201. Pubmed=9071574:
Chidambaram A., Gallani M., Gerrard B., Stewart C., Goldstein A.,
Chumakov I., Bale A.B., Dean M.;
"Characterization of a YAC contig containing the NBCCS locus and a
novel Kruppel-type zinc finger sequence on chromosome segment
q922.3."?
RT          922.3.?
CC          -1- FUNCTION: MAY BE INVOLVED IN TRANSCRIPTIONAL REGULATION.
            -1- SUBCELLULAR LOCATION: Nuclear (Potential).
            -1- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN KIDNEY, IS WEAKLY
            EXPRESSED IN HEART, LIVER, SPLEEN, AND SMALL INTESTINE, AND IS NOT
            EXPRESSED IN ADULT BRAIN OR SPINAL CORD.
            -1- SIMILARITY: BELONGS TO THE KRUPPEL FAMILY OF C2H2-TYPE ZINC-
            FINGER PROTEINS.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between The Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
EMBL: U08251. AAA70188.1; -.
HSSP: P25490. 10BD.
Gene: HGNC:12957; ZNF169.
MIM: 603404; -.

```

Query Match	Best Local Similarity	Score	DB 1:	Length	512:
Matches	159;	Conservative	66;	Mismatches	252;
				Indels	295;
				Gaps	28;
DR	InterPro: IPRO000823; znf.C2H2.				
DR	Pfam: PR00096; zf.C2H2.11.				
DR	SMART; SM00355; znf.C2H2; 10.				
DR	PROSITE; PS00028; ZINC_FINGER_C2H2_1; 9.				
DR	PROSITE; PS00157; ZINC_FINGER_C2H2_2; 12.				
KW	Transcription regulation; DNA-binding; Zinc-finger; Metal-binding;				
KW	Nuclear protein; Repeat.				
FT	NON_TER	1	1		
FT	DOMAIN	144	473	ZINC_FINGERS.	
FT	ZN_FING	144	166	C2H2-TYPE.	
FT	ZN_FING	172	194	C2H2-TYPE.	
FT	ZN_FING	200	222	C2H2-TYPE.	
FT	ZN_FING	228	250	C2H2-TYPE.	
FT	ZN_FING	311	333	C2H2-TYPE.	
FT	ZN_FING	339	361	C2H2-TYPE.	
FT	ZN_FING	367	389	C2H2-TYPE.	
FT	ZN_FING	395	417	C2H2-TYPE.	
FT	ZN_FING	449	473	C2H2-TYPE.	
FT	ZN_FING	479	504	C2H2-TYPE.	
SO	SEQUENCE	512 AA;	57651 MW;	58780DF763C63E9A2 CRC64;	
QY	Query Match	5.38;	Score 278.5;	DB 1:	Length 512;
Db	Best Local Similarity	20.68;	Pred. No. 1e-05;		
QY	Matches 159; Conservative	66;	Mismatches 252;	Indels 295;	Gaps 28;
Db	2	PAKFSPPGAASSSOLLKQYALSGHPTQI-----FPSSSAGD-----FOLEAPRCS	47		
QY	281	OHFFSVGVGVRSHKPTAPSPAL--PSTDLIASPHLAF-----GTGLGLAA	328		
Db	48	SEKGESE-----ETGPDSSLRKRPRIISITPFSPHGDDPVEMWEGNREGDLDLAAQ	100		
QY	329	--CLGAA-----RGLEAASPGLLKPKNGSGELGYGVISLSLEKRGGRHKCFKAKVFGSD	382		
Db	101	RMSLGSDDTLKAGDTSSEGAIVR--GNRYLGLRKK-SLSPSHOKHNVCPEGGRFCQR	156		
QY	383	SALQILHRSHTGERPKYKNCWGNPTTRTGNLKVHFRHREKRPYHGMNRPVPEHLDYLI	442		
Db	157	SDLIKHQRTHTGKRPYLCPEGGRFSQKAPSTIRHQKHSGER-----YVC	202		
QY	443	TSSG--LPGMSVPRPEKAEBAEATP-----GGVERKPLVASTALSAESTLLSTGT	494		
Db	203	RECGRRFRYTSLSLTHNRHISGERPFVQCEGRGRFRQK-----IAL-----LHQRT	249		
QY	495	STAAVAGLPTFNKPFVLMKVAEPKSKADNTPPGSEGSALAGVADSSATPMOLSKLYTSL	554		
Db	250	HLEERP-----FVCLISVEGALOKA-----	268		
QY	555	PSWALTNLHLKSTGSPFPFVYLEPLGASPSSETSKLOOLEIKDROGAVAAVASTAGAPT	614		
Db	269	----SLDQHQSSHHGERF-----LSLECGRTFRQDLSL--LSHQ-----	302		
QY	615	SAPAPSSSASGPNOCVLCRLVLSCPRAIRLHYGHGGERPFKCYKVGAFSTRGNIRAHF	674		
Db	303	----VHSGEKPRVCAECGSHFSROKVTILRHQRTHTGKRPYLCPCGGRFSQKYTL----	354		
QY	675	VGHKTPPAAARAQNSCPICQKKETNAVTALQOIVRMILGQIIRPGSALSSEGGAGAEJSSSE	734		
Db	355	IGHQRTHTGKRPYLCPCGGRFGGNGTILRHQRTHTG-----	391		
QY	735	OSTASPGSPPOPOOOPSPPEEEMSEDEEDEDVYTDDESLAGRSGSEGEKAISVR	794		
Db	392	-----EKYYLCKCGAGFGAFKSLLTR	412		
QY	795	GDSEEVSGAEEAVATVAATTVKEMDSNEKAPQHTLPPPPPPDNLDPHQPMEGTSDV	854		
Db	413	-----HQRTHS-----	418		
QY	855	SGAMEEAKLEGISPMWALTQOEGEGTSTPLVEELNLPAMKKDPDESSGRK--ACEVCG	912		
Db	419	----BEELIYVDRKCG-----QSLGQKSHLI-----SDQRTSHGEKCTIDCEG	457		

Db	1	IKSLIVLAESPVESELITTELVEKVIIVEKTAARVTOGSDRKQV-----	48
QY	238	PLFSPITKPAOTGKTAS-SSSSSSSSGAEPP-KQAFHLYHPIGSQHPFVSGVGRSHK	294
Db	49	-----PNOEGKEEAPVATAAQPKEDAEQPDAKEG-----PAEGQDP--GGVDNAAE	92
QY	295	PPPPAP-SPALPGSTDQILASPHLAFPECTTGLLAQCLGARGL-----E	337
Db	93	ASPAAPSPRPOPAEESEVGN--SSPEKGSDAPEST--EARGMELEGKEEAGEAVEDEE	147
QY	338	AAASPDLMPK-----NGSGELG-----YGEVTSLEKPGGRHKC	372
Db	148	EAKTPAAOPKSSKENAEDNEGSTDSGQENSETLLRSCTYSDRFEKRYAAVTHKC	207
QY	373	RECAKVFSGDSALQILHRSHTGERPYKCNVCGNPFTRKGLKVHHRHREKYPHVOMNH	432
Db	208	EEOGKFETHGNGKRRHRIHTGEKPFSCREGCKNAFSDPAACKAHBKTH-----	255
QY	433	PVEPHLDVYITSSGGLPYGMSVPPREKAEEEAECTPGGCVKRPVLVASTALATESILTST	492
Db	256	-----SPLKPYGC-----EECG-----KSYRLIS-274	
QY	493	GSTAVAPGLPTFNKPFYLMKAVEBKSKADEMTPTPGSESGALAGVADSGATRMQSLKVT	552
Db	275	-----LINLHKRRITGEKKYCD-----CGKLTFT	299
QY	553	SILPSMALTLNHLKSTGSPFPYVLEPLGAPSE-TSKLQOLV-----EKID	597
Db	300	T--SGMLKRRQQLVHSGE-KPYOCDYGRSPSDPTSKMRHLEHTDTEKHKCPHCCKFN	355
QY	598	ROG-----AVAVAS-----TASGAPITTSAPAPSSASGPRNOCYLCRLSCP	639
Db	356	QVONLKAHLKIHITADGPLKCREGNEFTTSG--NLKRLRIHSGEKPYYVCVHCQRQFADP	413
QY	640	RALRLHYGOGHGERPFCYKCGRAFSTRGNRAHE--VGHKTSPAARA-----	685
Db	414	GALQAHVPIHTEKPRCCCLGKFFTOASSLIANHRTDTEKPRPYCERCGRKFVSSQLA	473
QY	686	-----QNSCPLQCKRFTNAVTLQOAHRMILGQILPNGSALSEGGAAOENSSBO	735
Db	474	NHTRHNDNIRPHKCTYONKAFVNVGDSLKHIIINH-GEKPLCDRCGRGFNRVDNLRSHV	532
QY	736	STASGPGSPFPOPOSQDPREEEMSEEEDEDEEDVTDDESLAGSGSSGGEKAISVYG	795
Db	533	KTYH-----QGRAKMKILPEED-----GSEL--NIVTVAS	560
QY	796	DSEEVSGAEEVATSVAPITTKEMDSNEKAPQHTLPPPPRPDMLDHQPMEQG-TSPV	854
Db	561	DDMVTLATELALATAVQILTV-----PVAATAVTADE	592
QY	855	SGAMEEPAKLEGISSPMAALTOEGEGSTPLVEELINLEPMAMKDPGESSGRACEVCGOS	914
Db	593	TEALKRAEI-----TKAVQV-----QDAP-NTQILVACDSCGEK	626
QY	915	FPTOTALIEEHOKTH	928
Db	627	FLDATSLAQHVRH	640
RESULT 24			
ID	ZOT1_XENLA	STANDARD;	PRT; 898 AA.
AC	PI18751;		
DT	01-NOV-1990 (Rel. 16, Last sequence update)		
DT	01-OCT-1994 (Rel. 30, Last annotation update)		
DE	Oocyte zinc finger protein XI-COF.1 (Fragment).		
OS	Xenopus laevis (African clawed frog).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Amphibia; Batrachia; Anura; Mesobatrachia; Pipiloidea; Pipidae;		
OC	Xenopodinae; Xenopus.		
OX	NCBI_TaxID=8335;		


```

RN      [1]
RP      SEQUENCE OF 1-311 FROM N.A.
RX      MEDLINE=89345612; PubMed=2503827;
RA      Knoechel W., Poeling A., Koester M., el Baradi T., Nietfeld W.,
RT      Boumeester T., Pleier T.;
RL      "Evolutionary conserved modules associated with zinc fingers in
        Xenopus laevis.";
        Proc. Natl. Acad. Sci. U.S.A. 86:6097-6100(1989).
RN      [2]
RP      SEQUENCE OF 284-898 FROM N.A.
RX      MEDLINE=90040696; PubMed=2509712;
RA      Nietfeld W., El-Baradi T., Mentzel H., Pleier T., Koester M.,
RT      Poeling A., Knoechel W.;
RL      "Second-order repeats in Xenopus laevis finger proteins.";
        J. Mol. Biol. 208:639-659(1989).
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; M25866; AAA50013.1; .
DR      PIR; A33282; A33282.
DR      HSSP; P08047; 1SR2.
DR      InterPro; IPR000822; ZnF_C2H2.
DR      Pfam; PF00096; zf-C2H2; 21.
DR      SMART; SM00355; ZnF_C2H2; 21.
DR      PROSITE; PS00028; ZINC_FINGER_C2H2_1; 21.
DR      PROSITE; PS50157; ZINC_FINGER_C2H2_2; 21.
KM      ZINC-finger, Metal-binding; DNA-binding; Repeat.
FT      ZN_FING 289 311 C2H2-TYPE.
FT      ZN_FING 342 364 C2H2-TYPE.
FT      ZN_FING 370 392 C2H2-TYPE.
FT      ZN_FING 398 420 C2H2-TYPE.
FT      ZN_FING 425 448 C2H2-TYPE.
FT      ZN_FING 426 448 C2H2-TYPE.
FT      ZN_FING 454 476 C2H2-TYPE.
FT      ZN_FING 481 504 C2H2-TYPE.
FT      ZN_FING 482 504 C2H2-TYPE.
FT      ZN_FING 510 532 C2H2-TYPE.
FT      ZN_FING 537 561 C2H2-TYPE.
FT      ZN_FING 538 561 C2H2-TYPE.
FT      ZN_FING 567 589 C2H2-TYPE.
FT      ZN_FING 595 617 C2H2-TYPE.
FT      ZN_FING 623 645 C2H2-TYPE.
FT      ZN_FING 651 673 C2H2-TYPE.
FT      ZN_FING 679 701 C2H2-TYPE.
FT      ZN_FING 707 729 C2H2-TYPE.
FT      ZN_FING 735 758 C2H2-TYPE.
FT      ZN_FING 764 786 C2H2-TYPE.
FT      ZN_FING 792 814 C2H2-TYPE.
FT      ZN_FING 820 842 C2H2-TYPE.
FT      ZN_FING 848 870 C2H2-TYPE.
FT      ZN_FING 876 898 C2H2-TYPE.
FT      NON_TER 898
SO      SEQUENCE 898 AA; 101167 MW; 795D806E5696B0B6 CRC64;

Query Match 5.1%; Score 271.5; DB 1; Length 898;
Best Local Similarity 19.7%; Pred. No. 3.6e-05;
Matches 199; Conservative 95; Mismatches 387; Indels 327; Gaps 41;

```

```

QY      145 LGATPLPEESTPAPPPPPPPGVSGHLNIPL-----LLEELRVLQORIQHQ 196
           |||
DB      194 CSINPLTEIEIGDTPTP-----IMGCSLNTSLSDNYISDELKEEPTSCGRNOSDS 246
QY      197 MTEQICROYLLIGLQIVGAPASPELPGTGAASSTKPLPLPESP----- 242
           |||
DB      247 IT-----PLTEQIOGTDTPTPIMCCSLKDNKYDGNPHWSPKNTLRRKYSNCECH 295
QY      243 ---IKPAOTGKTTASSSSSSSGAEPKQAFHL---YHPLGSHPRSVGVGR----- 291
DB      296 EYLHKRDFGKHQMHKREKFSCEGKCFNLQCDPRHQTGK-PPSCSKGKCFAPL 354
QY      292 -----SHKPPAPSPALPGSTDOLIASPHLAFPTGTGLAAOCIGARGLLEAA 340
DB      355 SDLYNRRKHTGERPPRSCSGCGFTPRNMLIHNRHTTGKPRSCSEC----- 403
QY      341 SPGLLKPNGSELGYEYISLEKPRGRKRCFCAVFGSDALQILRSHTGERPYC 400
DB      404 --GKCFSSKQSS--LVNHQRTHIGEK--FCCSCDKCFASSSELINHQRTHTGKPRFC 456
QY      401 NVCGNRFTTNGNLKVFHNRHREKYPHYQMPHYRPHENDIYITSSGLPYGMVPRPKAE 460
DB      457 SECGRKCFTHMS---HFANHQ-----MIHQ----- 478
QY      461 EAGTPGGVERKPLVASTALSTESLTLLSTGTAVAPGLPTPNKFKVLMKAVEPKSKA 520
DB      479 -----EKRFPCSKGCKCFASSSDL-----TPHR-----RTHH 505
QY      521 DENTPPGSE-GSAIAGVADSGSATRMQL-----SKLVTSIPSMALLTNHLKSTGS 569
DB      506 REKTFSCSECGKCF---SNHSHLARHQMHITGKPRFCSECGKCFSSSGLLAHQORTIH 562
QY      570 PFPYVLEPLGASPSSETSKL--QQLVKKIDQGAVAVASTSGATTGAPARASSSASGN 627
DB      553 KYKPPSCSACGKCFSNRSHLIRHQMH-----TGKPP 595
QY      628 QCVILRLVLSCPRALRLHYGHHGGERPPKCVCGRAFTSGNLRHAFYHKTSPARAQN 687
DB      596 SCFECRKCFNSPNBLARHQHTHIGKPRFCSECGKCFASSDILTH--HRHTGSK-PF 651
QY      688 SCPIQKKTNAVTLQOHVBMILGQIIPNGSALSSEGGAAOENSSEOSTASGPGSFPDP 747
DB      652 SCSECGKCYSKSSLVHNQHTH--GEKPFSCSKCDKCFASSSELINHQHTHGEKAF-- 707
QY      748 QSQQSPSEEDMSEEBEDEEDVTDDEDSLARGSGSEGEKAISVGRGSE--VSG--- 802
DB      708 -----SCSECGKCFTHRSQLSRHQMHITGKPRISC-PPCECFVSSQL 750
QY      803 -AEEVATSVAAPTYKE-----MDSNKAPOHTLPPRPPPDNLDPQMEQTSVSGA 857
DB      751 TAAQOAHNMKVPFSCLECGKCFNSRNSFARHQMHITGKPRFSCSE---CRKGFSSQSL 807
QY      858 MEEKLEKGISSPMALTOEGEGTSTPLVERLNLPEAMKDPGESSGKRACEVGGSPPT 917
DB      808 ARHQ-----MTHNGE---KPF-----SCSECGKRFESN 831
QY      918 QTALENQKTHPRKDPGLFTCVPCNQGFLDKATLKKHMLAHNQVPPFA 965
DB      832 QSHLARHQMHITGKPR-FSCSECAKGFSSNOSGLARHQ-MTHNGEKPR 877

RESULT 25
ZFH1_DROME
ID ZFH1_DROME STANDARD; PRT: 1060 AA.
AC P28166;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Zinc-finger protein 1 (zinc-finger homeodomain protein 1).
GN ZFH-1.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachyera;

```



```

RESULT 26
ID Z084.XENLA STANDARD: PRT: 780 AA.
AC P18753;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 01-OCT-1994 (Rel. 30, Last annotation update)
DE Oocyte zinc finger protein XLOC8_4 (Fragment).
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipridae;
OC Xenopodinae; Xenopus.
NC NCBL_taxid=8355;
RN [1]
RN RN SEQUENCE OF 1-272 FROM N.A.
RX MEDLINE=89345612; PubMed=2503827;
RA Knoechel W., Poeling A., Koester M., el Baradi T., Nietfeld W.,
RA Boumeester T., Pleier T.;
RA Evolutionary conserved modules associated with zinc fingers in
RT Xenopus laevis.*;
RL Proc. Natl. Acad. Sci. U.S.A. 86:6097-6100(1989).
RN [2]
RN RN SEQUENCE OF 245-780 FROM N.A.
RX MEDLINE=90040698; PubMed=2509712;
RA Nietfeld W., El-Baradi T., Mentzel H., Pleier T., Koester M.,
RA Poeling A., Knoechel W.;
RL "Second-order repeats in Xenopus laevis finger proteins.";
RL J. Mol. Biol. 208:639-659(1989).
CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to licenset@isb-sib.ch).
CC -----
DR DR EMBL; M25868; AAA50015.1; -.
DR PIR; C33282; C33282.
DR PIR; S06548; S06548.
DR HSSP; P08047; ISP2.
DR Interpro: IPR000822: znF_C2H2.
DR Pfam: PF00096: zf-C2H2; 15.
DR SMART; SM00355; znF_C2H2; 15.
DR PROSITE; PS00028: ZINC_FINGER_C2H2_1; 15.
DR PROSITE; PS0157: ZINC_FINGER_C2H2_2; 15.
KW K zinc-finger; Metal-binding; DNA-binding; Repeat.
FT FT NON_TER 1 1
FT ZN_FING 250 272 C2H2-TYPE.
FT ZN_FING 278 300 C2H2-TYPE.
FT ZN_FING 306 328 C2H2-TYPE.
FT ZN_FING 334 356 C2H2-TYPE.
FT ZN_FING 362 384 C2H2-TYPE.
FT ZN_FING 390 412 C2H2-TYPE.
FT ZN_FING 418 440 C2H2-TYPE.
FT ZN_FING 446 468 C2H2-TYPE.
FT ZN_FING 474 496 C2H2-TYPE.
FT ZN_FING 618 640 C2H2-TYPE.
FT ZN_FING 646 668 C2H2-TYPE.
FT ZN_FING 674 696 C2H2-TYPE.
FT ZN_FING 702 724 C2H2-TYPE.
FT ZN_FING 730 752 C2H2-TYPE.
FT ZN_FING 758 780 C2H2-TYPE.
FT NON_TER 780 780
SQ SEQUENCE 780 AA; 88648 MW; 48E8C5BEC9B3915 CRC64;
Query Match 5.1%; Score 270.5; DB 1; Length 780;
Best Local Similarity 19.0%; Pred. No. 3.5e-05;
Matches 158; Conservative 93; Mismatches 301; Indels 279; Caps 30;
y 170 VCSGHINIPILIEELRVLQ-QROIHOMNTQICNVL-----LLGSIGGVIGA 217
:|||||: ||| :|::: |||: |||:

```

D	b	66	IRTHGVSIYFI	SDENMYITGN	KMLYEGIKEE	-PQQLRP	PCCEKEDKSDITAD	GGTLVN	1.22
Q	y	218	PASPELPGTGA	-ASSTKRLPL	DFSPK---	PAOTGKTTA	SSSSSSSGAEPR	KQAF	272
			: :	: :	: :	: :	: :		
D	b	125	YNEPSKIR	EGEDGFCAN	GNPTNP	SEISMPK	OPRANGIK	EEVASC	170
Q	y	273	HLVHPLG	SHPPSVG	VGRNKR	KTTPAS	-----	ALPGSTQ	LA
			: :	: :	: :	: :	: :	: :	
D	b	171	-----	GRSDCS	INPLED	IOGTDIP	TPIMGY	NHLIT	IQAIK
			: :	: :	: :	: :	: :	: :	
Q	y	320	GTYGGLA	AOCLGAR	GLEAAS	GLGLKPK	NSGSELG	YGEV	ISSLEK
			: :	: :	: :	: :	: :	: :	
D	b	225	HTTEKOF	FSHINIL	NHDL	STHAG	-----	KKP---	FG
			: :	: :	: :	: :	: :	: :	
Q	y	380	GSDSAL	QLIHLR	SHHTGER	PRYKCN	OGNFR	TTGNL	KLVHFR
			: :	: :	: :	: :	: :	: :	
D	b	260	ASSSELN	VHRRT	HTRK	PFCS	QCGKCF	SNQTK	LKLH
			: :	: :	: :	: :	: :	: :	
Q	y	440	YVITSS	GLPYG	MSVPR	PEKAE	EEACT	PGCGV	ERKPLV
			: :	: :	: :	: :	: :	: :	
D	b	306	-----	-----	-----	-----	-----	-----	-----
Q	y	500	PGLPTNK	EVLMK	AVEPK	SKA	DENTP	PGSE	-GSA
			: :	: :	: :	: :	: :	: :	
D	b	320	-----	-----	-----	-----	-----	-----	-----
Q	y	559	LITNLH	LKSG	FPFPV	LEPLG	ASPS	SETSK	LQOL
			: :	: :	: :	: :	: :	: :	
D	b	349	-VTVH	-RTH	TGKPR	YPC	SCQCGK	CFTR	SSDL
			: :	: :	: :	: :	: :	: :	
Q	y	619	PSSSAS	GPNO	CVLC	RLSP	ALRL	LYG	ONG
			: :	: :	: :	: :	: :	: :	
D	b	383	-HTG	KRPS	CSHG	KFTTS	SELNV	HRHTG	ERKPR
			: :	: :	: :	: :	: :	: :	
Q	y	679	TS	PARA	ON	SC	PC	QK	KFTNA
			: :	: :	: :	: :	: :	: :	
D	b	442	EEKP	-----	SCV	CGK	CF	SKD	HLK
			: :	: :	: :	: :	: :	: :	
Q	y	734	---	EOS	TAS	GP	SP	PO	QSO
			: :	: :	: :	: :	: :	: :	
D	b	494	KI	HR	AD	FC	SK	NLT	NP
			: :	: :	: :	: :	: :	: :	
Q	y	790	AT	SV	R	D	S	E	V
			: :	: :	: :	: :	: :	: :	
D	b	539	D	Y	S	I	N	L	T
			: :	: :	: :	: :	: :	: :	
Q	y	850	G	T	S	V	A	G	M
			: :	: :	: :	: :	: :	: :	
D	b	569	N	K	D	V	A	N	A
			: :	: :	: :	: :	: :	: :	
Q	y	905	R	K	-	A	C	E	V
			: :	: :	: :	: :	: :	: :	
D	b	615	D	K	L	S	E	C	E
			: :	: :	: :	: :	: :	: :	

Db 326 RDLDATID---ISPKOSTHGERG-----HRCSDCGKFFLQASNFIQHRRIH 368

QY 393 TGERPYACNCGNFTTTRGNLKVPHRHRERKXYH-----YQMNPHVPRHLDVYITS 444

Db 369 TGEKPRFCGEGCKSYNQNVHLTHQRVHTGKERKYQOVCQKAFRVASHLYQNH---SVHS 425

QY 445 SGDPYKGSVYPERKAEKEEGTGGGVERKRYLAVSTALSTESTILLSTGTSTVAPRLPT 504

Db 426 GERPYGCN-----EKGKMG--RHSHLI-----EHLKRRHFRKESQRCSDKRSK 466

QY 505 ENKFVLEKAVEPKSKADENTPPGSEGSALAGVADSGSATRMOQLSVTLSPWALLTNHL 564

Db 467 NTMLSVAKKLTSEYSEAD-----MELS-----487

QY 565 KSTGSEFPFVYLPRLGASPSSETSKIQQLVE-----KIDRGAVAVASTAGAPPTS--- 615

Db 488 -----GKTQRVNSQVYDGECEGEPQGLDKRKGIPMKELL-GPPSKRMN 531

QY 616 -ADAP-----SSASGPNOCVITCLRVLSCPRALRLHYGNGRPERPKCKVCGRAFSTRGN 669

Db 532 YSEVPYVHKKSSSTERRHKCKECSKFIOQSAHLIOQRHHTGKPRCECGKSY---N 587

QY 670 LRAHVEGHKTSPPARAONSCDPCQKFFTNAYVTLQOHVHMHLGQILPMGSSALSEGGGAQ 729

Db 588 ORVHLTHQHVHTGTEKRYTCLCGKAFRVRSHTLYOHSVH--SGERPFKCNCKGCGRRS 646

QY 730 ENSSEGSTAGSPGSPFPQSQSPSPDEEMSEFEDEEEDVETDDSLAGRSSESGEK 789

Db 647 HLGHLRLHS-----REKSHOCCRGELFQYVSLI-----EH 679

QY 790 AISRQDSEEVSGAESEVATVSAAPTVEKENDSNKAPQHTLPPPPPDNLDPQPMEQ 849

Db 680 OVLHMGOQKNKNGICEAVYSWNLTVIEDKLTLEQ-----QPYQ- 718

QY 850 GTSVDYSGAMEEELKLEISPMALLTDBEGTSTPLVEELNPEAMKRPDESSGK--- 906

Db 719 --CDIG-----KAFYSDDL-----QHYRHTAKRPYQCDICRENVGQCSHTKOHQ 764

QY 907 -----ACEVCGSGFPOTALEEHOKTPRKDPLFTFCVCGROGFLDRATLTKHMLA 957

Db 765 KLYSPTKSHQCHCGRGFTLTKSHLNGHRIHTGKPR-FQCKRCGNMFSWCSILFKH-LRS 822

QY 958 HHQVPPAPRHGPONIALTSL 977

Db 823 HERTDP-----INTLSV 834

RESULT 28

ZN45_HUMAN STANDARD; PRT; 682 AA.

ID ZN45_HUMAN

AC 002386; P78472.

DT 01-OCT-1993 (Rel. 27, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE zinc finger protein 45 (BRC1744).

OS ZNF45.

CS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Catarrhata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Canarihini; Hominoidea; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

FX MEDLINE=97220056; PubMed=9067431;

RA Constantinou-Deltas C.D., Bashlades E., Patsalis P.C.,

RA Hadjimarco M., Kroisel P.M., Ioannou P.A., Roses A.D., Lee J.E.;

RT "Complete coding sequence, exon/intron arrangement and chromosome

RT location of ZNF45, a KRAB-domain-containing gene.";

RL Cytogenet. Cell Genet. 75:230-233(1996).

RN [2]

RP SEQUENCE OF 1-400 FROM N.A.

RP MEDLINE=92217982; PubMed=1559709;

RA Constantinou-Deltas C.D., Gilbert J., Bartlett R.J., Herdstreith M.,

RA Roses A.D., Lee J.E.;


```

RT      "The identification and characterization of KRAB-domain-containing
RT      zinc finger proteins."
RL      Genomics 12:581-589(1992).
CC      -1- FUNCTION: MAY BE INVOLVED IN TRANSCRIPTIONAL REGULATION.
CC      -1- SUBCELLULAR LOCATION: Nuclear.
CC      -1- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-
CC      FINGER PROTEINS.
CC      -1- SIMILARITY: CONTAINS 1 KRAB DOMAIN.
CC      -----
CC      This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb.ch/announce/
CC      or send an email to license@isb.slb.ch).
CC      -----
DR      EMBL; L75847; AAB05653.1; -.
DR      EMBL; M67509; AAA36133.1; -.
DR      EMBL; M67509; AAA36134.1; -.
DR      HSSP; P08046; 1A11.
DR      TRANSFAC; T04988; -.
DR      GeneW; HGNC:13111; ZNF45.
DR      MIM; 601781; -.
DR      InterPro; IPR001909; KRAB.
DR      InterPro; IPR000822; ZnF_C2H2.
DR      Pfam; PF00096; zf-C2H2; 15.
DR      Pfam; PF01352; KRAB; 1.
DR      PRINTS; PR00048; ZINC_FINGER.
DR      PRODOM; PD000003; ZnF_C2H2; 13.
DR      SMART; SM00349; KRAB; 1.
DR      SMART; SM00355; ZnF_C2H2; 15.
DR      PROSITE; PS50805; KRAB; 1.
DR      PROSITE; PS00028; ZINC_FINGER_C2H2_1; 15.
DR      PROSITE; PS50157; ZINC_FINGER_C2H2_2; 18.
DR      Transcription regulation; DNA binding; Zinc-finger; Metal-binding;
DR      Nuclear protein; Repeat; Polymorphism.
KW      DOMAIN
KW      8
KW      78
KW      KRAB.
KW      ZINC_FINGERS.
FT      DOMAIN
FT      ZN_FING 192 662
FT      ZN_FING 192 214
FT      ZN_FING 276 298
FT      ZN_FING 304 326
FT      ZN_FING 332 354
FT      ZN_FING 360 382
FT      ZN_FING 388 410
FT      ZN_FING 416 438
FT      ZN_FING 444 466
FT      ZN_FING 472 494
FT      ZN_FING 500 522
FT      ZN_FING 528 550
FT      ZN_FING 556 578
FT      ZN_FING 584 606
FT      ZN_FING 612 634
FT      ZN_FING 640 662
FT      VARIANT 187 187
FT      A -> T (IN DBSNP:1047452).
FT      /FTId=VAR_012019.
FT      R -> K (IN DBSNP:399098).
FT      /FTId=VAR_012020.
FT      T -> A (IN DBSNP:388706).
FT      /FTId=VAR_012021.
FT      P -> R (IN DBSNP:388685).
FT      /FTId=VAR_012022.
FT      R -> K (IN DBSNP:407731).
FT      /FTId=VAR_012023.
SQ      SEQUENCE 682 AA; 78242 MW; 9C8E536EB00FCE8B CRC64;
Query Match 5.18; Score 270; DB 1; Length 682;
Best Local Similarity 22.08; Pred. No. 3.2e-05;
Matches 132; Conservative 56; Mismatches 189; Indels 222; Gaps 23;

```

```

Yy 442 EKKYHQMHPHYREHLDYVITS--SGIPYQMSV-----PEKAEEGAHPGGVEKPLV 475
Db 218 KSYTN-----DASTSFSSORSHLPHHQRVPTGENPKYVEECGRNVKGS----- 260
Qy 476 ASTALSAEESLTLSTGTSTAVAPGLPTFNKFLVLMKAVEPKSKADENTPPGSEGAIAIG 535
Db 261 -----SHQAPLIVHTG-----EKRYKCE-----CG 282
Qy 536 VADSGATPMQSLKLTSLPSMALLTNHLK-STGSGFPPTYVLEPLGASSETSLOOLVE 594
Db 283 V--GFSQR-----XYQLWHLKVHTGK--PKYCEECGKSFMSWRSLQ----- 320
Qy 595 KIDRGAAVAVASTASGAPTTASAPSSASASPNOCVICRLVSCPRALRYHYQHGGER 654
Db 321 -----AHERIHTEKPKYKMACGKSFYSYSHLNIHCRITHGEKP 359
Qy 655 FKCKVCGAFSTRTGNLRAHFVGHKTPSPARAQNSCPICQKFTNAVTLQOHVNMHLGQI 714
Db 360 YKCECGKGGKGSVSHLQAHQISH---TGKPKYKECGKGCRASNLLDHQHGT--GEK 414
Qy 715 PNGGSALEBGGGAQENSSEOSTYASGCPSPFOPQSOQSPEEDMSDEEEDDEEDVD 774
Db 415 PYQCDACGKGFSRSPDNIFRHVHTG-----EKPKCEE----- 448
Qy 775 EDSLAGRSGSGGGEKALTSVDSGEVSGAAEEVATSVAAPTYKEMDSNEKAPQHTLPP 834
Db 449 ---CGKGFSQASNLNLAHQKHTE----- 469
Qy 835 PPPDNLDPHPQMEQGTSDVSGAMEEELKLBGISSPMALTOEGEGTSPPLVELNPEA 894
Db 470 -----KPKKCGTC-----GKGFSSRS--SPLNV--H 490
Qy 895 MKKDGESSGRKACCEVCGSFPPTOTALFEHQKTHPKDGLFCVCFRQGLDNATLKKH 953
Db 491 CRHHTGEKPKK--CERCCKAFSPQSFSLQVHQRVHTGEKP--YQCAECGKGSFVSGQLQAH 546

RESULT 29
Evil_MOUSE STANDARD: PRT: 1042 AA.
AC P14404;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Ecotropic virus integration 1 site protein.
GN Evil OR EVI-1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88311086; PubMed=2842066;
RA Morishita K., Parker D.S., Mucenski M.L., Jenkins N.A., Copeland N.G.,
RA Ihle J.N.;
RT "Retroviral activation of a novel gene encoding a zinc finger protein
RT in IL-3-dependent myeloid leukemia cell lines.";
RL Cell 54:831-840(1988).
CC -! SUBUNIT: MAY INTERACT WITH CTBP1 (BY SIMILARITY).
CC -! SUBCELLULAR LOCATION: Nucleolus.
CC -! CAUTION: IT IS UNCERTAIN WHETHER MET-1 OR MET-1L IS THE INITIATOR.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb.ch).
CC -----
CC EMBL; M21829; AAA40581.1; ALT_INIT.
DR PIR; A31591; A31591.

```


R HSSP; P08047; 1SP2.
R TRANSEAC; T00273; .
R MGD; MGI:95457; Ev11.
R InterPro: IPR000822; ZnF_C2H2.
R Pfam: PF00096; zf-C2H2; 10.
R PRINTS; PR00048; ZINC_FINGER.
R ProDom; PD000003; ZnF_C2H2; 4.
R SMART; SM00355; ZnF_C2H2; 10.
R PROSITE; PS00028; ZINC_FINGER_C2H2_1; 8.
R PROSITE; PS0157; ZINC_FINGER_C2H2_2; 10.
R Zinc-finger; Metal-binding; DNA-binding; Developmental protein;
R Repeat; Proto-oncogene; Chromosomal translocation.
T DOMAIN 21 239 ZINC_FINGERS 1.
T DOMAIN 421 434 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
T DOMAIN 724 803 ZINC_FINGERS 2.
T DOMAIN 877 928 ASP/GLU-RICH (ACIDIC).
T ZN_FING 21 44 C2H2-TYPE.
T ZN_FING 75 97 C2H2-TYPE.
T ZN_FING 103 125 C2H2-TYPE.
T ZN_FING 131 154 C2H2-TYPE.
T ZN_FING 160 182 C2H2-TYPE.
T ZN_FING 188 210 C2H2-TYPE.
T ZN_FING 217 239 C2H2-TYPE.
T ZN_FING 274 296 C2H2-TYPE.
T ZN_FING 752 775 C2H2-TYPE.
T ZN_FING 781 803 C2H2-TYPE.
T SITE 553 557 CTBP-BINDING MOTIF 1 (BY SIMILARITY).
T SITE 584 588 CTBP-BINDING MOTIF 2 (BY SIMILARITY).
T SEQUENCE 1042 AA; 116847 MW; 8DEDF164F536D2FE CRC64;

Query Match 5.1%; Score 269; DB 1; Length 1042;

Best Local Similarity 21.9%; Pred. No. 5.4e-05;

Matches 162; Conservative 72; Mismatches 283; Indels 224; Gaps 31;

367 GGRHCRFCAKYGSGSALQIHLS-ITGERPYKMGNGNRRTTRGNLKVHFRHH----- 420
128 GKHTCEKCAKVFTRPSNLRHRSOHVAGAHACRECGKTRTSGGLKQHKHHSVKR 187
421 -----REKYPHV-----QMPH-PVPEHLDVIT 443
188 FICEVCHKSYQFESNCRHMYADCRQIKCKDCGMESTSLKHKHRCCKKNFPA 247
444 SSGLYGMVPRPKAEDEA-----GTPG-----GGEKPLVASTALSATSLTSLT 492
248 GGFEGQGGISLPGPRADKTSVMVNSHANPGADYFGTNRHP-----A 289
493 GTSYAVAPCLPTFNKFLVMAKAVEBKSKADENTPRSGESALIGVADSGSATRMQLSKLVT 552
290 GLTPRTAAGFSFSGFLFESGLNHRPLIRASPP-----VGLSTEGSNKCO-SPLLT 342
553 SLPSWALLTNHLSKSTGSPRRYVLEPLGASPSSTSKLOLVKEDRQGA---VAVASTAS 609
343 HPQLRAPQDILKALSKHNPVGNKRVELLPERSE-ERPLKISQSESSDDLDVSTPS 401
610 GA--PTTAPARASSASGRQCVCYCLAVLSCPRRLRLHNGHGERPRFKCKVCGRATSTR 667
402 GSDLETTSGSDLESDEDE-----CKENGKMKR-- 432
668 GNLRHNFHGHKTSAPARAONSCPLCKK-----FTNAVTLQOHVHMHGGLDIPNGSGA 720
433 -----DKVSP---LQVLASTNKKENHNHSYFASVEBQSAV---SGAVNDSIKA 476
721 LSE-----GGGAOENSSSEOSTASGPGSPROP-----OSQOOPSE----- 755
477 IASIAEKYFGSGVLQDKKVGALPRFSMFLPLFPFPAFSQSMYPPRPDLRLPLKMP 536
756 ---EEMSEEEDEDEEDVT---DEDSLARGSESGEKAIS-----VRGDSSEVS 801
537 QSPSEVKKLQKSSSESPDLTKRKDEKPLTSGPSKPGTATSDQOPLDLSMGSRGAS 596
802 GAE--EEVATSYAAPITVAKEMD---SNEKAPQHILPP----- 834
597 GTKLLEPRKNHVFGEKKGSMDTRPSSDGLQHARPTPEFMPIYVVERKRLDPLRLALK 656

QY 835 ---PPPDNLDPHPMEQGTSPVSGAMEEBA-KLEGISS--PMAA-LTQEGEGTSPPLV 886
DB 657 EKYLRPSGFLFHPQ-----MSAIEENAEKLESALKPEASELQSVSMFSFRA 707
QY 887 EELNLEPAKKDPGESSGRRACEVGSGFPOTALEEHQKTHPKDGLFTVCRCRGELD 946
DB 708 PPNTLPENLIRKGE---RYTCRCYCGKIFPRSANLFRHLRHTTGEP-YRCKYCDRSFSI 763
QY 947 RATLKKHMLAHQVPPRPAPH 967
DB 764 SSNLDQHVANNHKKERFKCH 784

RESULT 30

PRD_HUMAN STANDARD; PRT; 717 AA.
ID PRD_HUMAN
AC Q9H4Q3;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE PR-domain zinc finger protein 13.
GN PRDM13 OR PFM10.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA yang X.-H., Huang S.;
RT "A family of novel PR-domain (PRDM) genes as candidate tumor suppressors".
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
CC - FUNCTION: MAY FUNCTION AS A TRANSCRIPTION FACTOR.
CC - SUBCELLULAR LOCATION: Nuclear (potential).
CC - SIMILARITY: CONTAINS 1 SET DOMAIN.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by, and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AY004253; AAG13448.1; .
DR Genew; HGNC:13998; PRDM13.
DR InterPro; IPR001214; SET.
DR InterPro; IPR000822; ZnF_C2H2.
DR Pfam; PF00096; zf-C2H2; 4.
DR SMART; SM00317; SET; 1.
DR SMART; SM00355; ZnF_C2H2; 4.
DR PROSITE; PS0280; SET; 1.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 4.
DR PROSITE; PS0157; ZINC_FINGER_C2H2_2; 4.
KW Transcription regulation; DNA-binding; Zinc-finger; Metal-binding;
KW Nuclear protein; Repeat.
FT DOMAIN 1 126 SET.
FT ZN_FING 147 169 C2H2-TYPE.
FT ZN_FING 583 605 C2H2-TYPE.
FT ZN_FING 611 633 C2H2-TYPE.
FT ZN_FING 640 663 C2H2-TYPE.
FT DOMAIN 311 317 POLY-ALA.
FT DOMAIN 362 373 POLY-HIS.
FT DOMAIN 382 388 POLY-PRO.
FT DOMAIN 564 575 POLY-GLY.
SQ SEQUENCE 717 AA; 75059 MW; FB45135F6488254A CRC64;
Query Match 5.1%; Score 268.5; DB 1; Length 717;
Best Local Similarity 24.5%; Pred. No. 3.9e-05;
Matches 173; Conservative 55; Mismatches 262; Indels 215; Gaps 34;
QY 132 AGGGGGIL-----ASPKLGATPLPESTPAPP-----PPPPPPGVSAG 173

Db 173 SGGGGAGFLHNEHARAGAVPADGLGLSPKRPAPDPAPASQACTLRPHRLGPP----- 227
Qy 174 HLNLPLLEELRVLDQROHQOMTEQICROVLLGLSGTVGARPSASELPGTAASST 233
Db 228 -----VQAC-----GAREGJKREASSAP-----SAT 248
Qy 234 KPLRLPLSPKPAOTGKTATSSSSSSSSSGAERPKQAFPHLYHPLGSHQHPFVGGVGRSH 293
Db 249 SP-----TPGKWQKKKKGEQDLRALDMSGARGQHFLGI-----VGG----- 287
Qy 294 KPTPAPSPALPGSTQDLIASPHLAF-PGTTGLLAQCLGARGLEAASPGILKPKNGSG 352
Db 288 -----SSAGVS-----LAFYPGVRS-AFKPAGLARA-AAAHQDPYHESSS 328
Qy 353 ELGYGEVYSSLEKPPCGRHKCRCAKVFSGDSALQIHLRSHTEGPPYKCNVCNRTTTCGN 412
Db 329 KQAGLALGRL--LGGGRACG-----RPGSGEN-----SAAAG 359
Qy 413 LKVPFHRREKYPHYOM-----NPHVPEHLDYITSSGLPYGMSVPEKAEEAGTPGG 467
Db 360 AGHHHHHHHHHHKCLLAGDRPPPPPELPCSGALRGFPL-LSVPEEA-----SAFK 413
Qy 468 GVERKPLVASTTALSTESLTLSTGTSTAVABQLPTFNKFLMKAVEPKSKADENTPPG 527
Db 414 HVERAPRPAAL-ALPGARVAQL-----PRAPGLP-----LERCALPRLPDGGLKATPG 459
Qy 528 SEGSAIAGVADSGSATRMQLSKLVTSPLPSMALLTNHLKSGSPFP-----VYLEPLGA-- 581
Db 460 GECSHLPLVMPAFVYNGEL--LYGSPATAYVPLKLHFGGLVLPESISYFSGPAAAL 517
Qy 582 SPSEFSKL-----QOLVE-----KIDROGAVAVASTASAPTSAPAPSS 622
Db 518 SPALGSLASIDRELTAMHQOLSEMAAGKGRGLD-SGLPRAVAAGG-TGGGSGSG 575
Qy 623 ASGP--NQCVCILRVLSCPRALRLHYGQHGGERPKCKYCGRAFSTGPNLRAHFVGHKT 679
Db 576 AGKPKTGHCLCYGKLYSKYGLKIMRHTGYKPLCKCKVCLRPGRDPSNLNKHRLH-- 633
Qy 680 SPARAQNSCPLCQKFTNAVTLQOVVRMHLGGQIPNGSALSSEGGGAQENSSEOSTAS 739
Db 634 -AEGNTPYRCFCGKYLVRRLDERHVKSRHPCQ-----SLIARAG-----D 674
Qy 740 GPGSPPOSOQOPSPREESEEEDEEDVDEEDSLAGSGE 784
Db 675 GPGAP-----GYPPEQDPKSDSDVVCFTDDGSDPE--VGGGE 714
RESULT 31
Y441_HUMAN STANDARD: PRT: 697 AA.
ID Y441_HUMAN
AC 043167;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical zinc finger protein KIAA0441.
CN KIAA0441.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
MEDLINE=98116655; Pubmed=9455477;
RA Ishikawa K.-I., Nagase T., Nakajima D., Seki N., Ohita M.,
RA Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
RA "Prediction of the coding sequences of unidentified human genes. VIII.
RA 78 new cDNA clones from brain which code for large proteins in
RA vitro.";
RL DNA Res. 4:307-313(1997).
CC -1- FUNCTION: MAY BE INVOLVED IN TRANSCRIPTIONAL REGULATION.
CC -1- SUBCELLULAR LOCATION: Nuclear (potential).

CC -1- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-
CC FINGER PROTEINS.
CC -1- SIMILARITY: CONTAINS 1 BTB/POZ DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>
CC or send an email to license@isb-sdb.ch).
CC -----
DR EMBL: AB007901; BAA23713.1; -
DR HSSP: P08046; 1AIG.
DR InterPro: IPR000637; AT_hook.
DR InterPro: IPR000210; BTB_POZ.
DR InterPro: IPR000622; ZnF_C2H2.
DR Pfam: PF00096; zf-C2H2; 8.
DR Pfam: PF00651; BTB; 1.
DR Pfam: PF02178; AT_hook; 1.
DR PRINTS: PR00048; ZINC_FINGER.
DR PRODOM: PD000003; ZnF_C2H2; 2.
DR SMART: SM00384; AT_hook; 1.
DR SMART: SM00225; BTB; 1.
DR SMART: SM00355; ZnF_C2H2; 8.
DR PROSITE: PS00097; BTB; 1.
DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; 8.
DR PROSITE: PS0157; ZINC_FINGER_C2H2_2; 8.
KW Hypothetical protein; Transcription regulation; DNA-binding;
KW Zinc-finger; Metal-binding; Nuclear protein; Repeat.
FT DOMAIN 10 133
FT ZN_FING 294 512 ZINC_FINGERS.
FT ZN_FING 294 316 C2H2-TYPE.
FT ZN_FING 322 344 C2H2-TYPE.
FT ZN_FING 350 372 C2H2-TYPE.
FT ZN_FING 378 400 C2H2-TYPE.
FT ZN_FING 406 428 C2H2-TYPE.
FT ZN_FING 434 456 C2H2-TYPE.
FT ZN_FING 462 484 C2H2-TYPE.
FT ZN_FING 490 512 C2H2-TYPE.
SQ SEQUENCE 697 AA; 78292 MW; F2BD3C144626544 CRC64;
Query Match 5.1%; Score 268; DB 1; Length 697;
Best Local Similarity 21.0%; Pred. No. 4e-05;
Matches 115; Conservative 68; Mismatches 169; Indels 196; Gaps 20;
Qy 364 EKPQG-RHKCRFCAGKVFSGDSALQIHLRSHTEGPPYKCNVCNRTTTCGNLVHFRHRE 422
Db 287 KRPGPEARCKDCGKVFYKYNHFLAIHQRSHTGERPFCNCGKGPAAOKSHLSIQVHTHTG 346
Qy 423 KYPHYQMNHPVPEHNDYVITSSGLPYGMSVPEKAEEAGTPGGVEKRPVASTALS 482
Db 347 ERPYT-----CTVCSKAL-----TTKHS 364
Qy 483 ARESLTLSTGTSTAVABGLPTF-----NKFVLMKAVEPKSKADENTPPGSEGSALAGVA 537
Db 365 LLEHMSLHS-----GKSFCTDCGKTFQSQR-QLKSHYRHT-----CHSLPECK 409
Qy 538 DSGSATRMQLSKLVTSPLPSMALLTNHLKS-TGSPFPVYLEPLGASPSSETSKLQLVETK 596
Db 410 DC-HRKFMDVSO-----LKKHLRTHGKEKPF----- 434
Qy 597 DRQGAVAVASTASGAPTSAPAPSSASGPNQCVCILRVLSCPRALRLHYGQHGGERPK 656
Db 435 -----TCEICGKSFYAKSSLSQTHIRIRHGEKPY 463
Qy 657 KAVCGRAFSTRGNLRAHFVGHKTSPARAQNSCPICOKKFTNAVTLQOVVRMHLGGQIPN 716
Db 464 CQICGKSPSDSSAKRRKCIILH-----TGKKPSCPCNQLQFALDNLKALKIKHSREKHS 519
Qy 717 GSAISEGGGAQE-----NSSEOS-----TA 738
Db 520 DASSIS-GSSNTEVEVRNILOLPYQLSTSGEQEIQLVTDVSHNINFMPPGSGISIVFA 578


```

Y      739 SGPSFPQP-----SQSPPEEM-SEEEDEDEEDVDDEDLAGSGSSGGRKA    790
Y      579 ESSONMTADQANLTLTQQPPOLOLLISAOEQETHTIOSLMNIESQM-PQSOTPVHV    637
Y      791 ISVGDSEBEVSAGEE-----VATSVAAPTTYKENDSMKEAFQHTLPAPPPEPDNDHP    844
Y      11: : : | : | : : : : : : : : : : : : : : : : : : : : : : : :
Y      638 ITLSKETLEHLAHQEQLREELHATSTSDPAOHLQLTQE-----PCPPPHNVHPQF    689
Y      845 QPMEOGS 852
Y      690 TPICQEQS 697
B
RESULT 32
D192_HUMAN
Z192_HUMAN STANDARD; PRT; 578 AA.
Q15776; Q9HAT1;
16-OCT-2001 (Rel. 40, Created)
16-OCT-2001 (Rel. 40, Last sequence update)
15-JUN-2002 (Rel. 41, Last annotation update)
Zinc finger protein 192 (LD5-1).
ZFPI92.
Homo sapiens (Human).
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NCBI_TaxID=9606;
[1]
SEQUENCE FROM N.A., AND VARIANT IEU-163.
TISSUE=Ovary;
MEDLINE=97386587; PubMed=9244436;
Lee P.L., Gelbart T., West C., Adams M., Blackstone R., Beutler E.;
"Three genes encoding zinc finger proteins on human chromosome 6p21.3:
members of a new subclass of the Kruppel gene family containing the
conserved SCAN box domain.";
Genomics 43:191-201(1997).
[2]
SEQUENCE FROM N.A.
Williams S.;
Submitted (SEP-2001) to the EMBL/GenBank/DDBB databases.
[3]
PARTIAL SEQUENCE FROM N.A.
TISSUE=Ovary;
MEDLINE=96230927; PubMed=8673473;
Beutler E., Gelbart T., West C., Kuhl W., Lee P.;
"A strategy for cloning the hereditary hemochromatosis gene.";
Blood Cells Mol. Dis. 21:206-216(1995).
-1- FUNCTION: MAY BE INVOLVED IN TRANSCRIPTIONAL REGULATION.
-1- SUBCELLULAR LOCATION: Nuclear (Potential).
-1- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-
FINGER PROTEINS.
-1- SIMILARITY: CONTAINS 1 KRAB DOMAIN.
-1- SIMILARITY: CONTAINS 1 SCAN BOX.
-----
THIS SWISS-PROT ENTRY IS COPYRIGHT. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see http://www.isb-sib.ch/announce/
or send an email to license@lsb-sib.ch).
-----
EMBL: U5796; AAB02260.1; -.
EMBL: AL358933; CA051901.1; -.
EMBL: U88080; AAC51656.1; -.
EMBL: U88079; AAC51656.1; JOINED.
HSSP: P07248; IPAA.
GeneW: HGNC:12983; ZNF192.
MIM: 602240; -.
InterPro: IPR001909: KRAB.
InterPro: IPR003309: Treg_SCAN.
InterPro: IPR000822; ZnF_C2H2.
Pfam: PF00096; zf-C2H2; 9.
```

DR Pfam: PF01352; KRAB: 1.
 DR Pfam: PF02023; SCAN: 1.
 DR PRINTS: PRO0046; ZINC_FINGER.
 DR ProDom: PD000003; ZnF_C2H2: 8.
 DR SMART: SM00349; KRAB: 1.
 DR SMART: SM00431; LER: 1.
 DR SMART: SM00355; ZnF_C2H2: 9.
 DR PROSITE: PS50805; KRAB: 1.
 DR PROSITE: PS50804; SCAN_BOX: 1.
 DR PROSITE: PS50028; ZINC_FINGER_C2H2_1: 9.
 DR PROSITE: PS50157; ZINC_FINGER_C2H2_2: 9.
 DR Transcription regulation; zinc-finger; Metal-binding; DNA-binding;
 KW Nuclear protein; Repeat; Polymorphism.
 FT DOMAIN 51 133 SCAN_BOX.
 FT DOKAIN 220 316 KRAB.
 FT DOMAIN 322 568 ZINC_FINGERS.
 FT ZN_FING 322 344 C2H2-TYPE.
 FT ZN_FING 350 372 C2H2-TYPE.
 FT ZN_FING 378 400 C2H2-TYPE.
 FT ZN_FING 406 428 C2H2-TYPE.
 FT ZN_FING 434 456 C2H2-TYPE.
 FT ZN_FING 462 484 C2H2-TYPE.
 FT ZN_FING 490 512 C2H2-TYPE.
 FT ZN_FING 518 540 C2H2-TYPE.
 FT ZN_FING 546 568 C2H2-TYPE.
 FT VARIANT 163 163 P->L.
 FT /FTID=VAR_009877.
 FT CONFLICT 34 34 A->D (IN REF. 2).
 FT CONFLICT 113 113 D->E (IN REF. 2).
 SQ SEQUENCE 578 AA; 65757 MW; 5914F5439F4B8AB6 CRC64;

Query Match 5.0%; Score 267; DB 1; Length 578;
 Best Local Similarity 22.0%; Pred. No. 3.7e-05;
 Matches 131; Conservative 56; Mismatches 195; Indels 214; Gaps 19;

```

QY 177 IPILIELRYLQOROHOMOMTFOIC-----ROVLIGS-----LGOTVGAP 218
Db 100 LILPELQTLVND--HOLENGEEVLTLEDLERIDILIGRPVSAYHGRVLMEEVYHS 157
QY 219 ASPSELPGT--GAASSTKRLPLPSPIKPAOTGKTTASSSSSS-----SSG-- 263
Db 158 ASAPERPNTQLEATQHKSVPQESQGRAMSTSQSPTRSQKSGSGDQEMTATLTITAGFQ 217
QY 264 -----AEPKQAFHLLPYLHPSQHPSPVSGVGSGSHKTPAPSPALPG 305
Db 218 TLEKIDMAVSLIREWLLDPQKDLCDRNREPNRMFSLGGETSENNELA----- 270
QY 306 STDOLIAS---PHLAPRGTTGLLAOCIG-AARGLEAASPGLLPKPKNGSGELGCEVIS 361
Db 271 -SKQVLTSTGIQPH-----GETAKKCGDYIRGLDHEEARDLL-----G 307
QY 362 SLEKPGG-----RHKCRFCAKYVGSDSALQILHRSHTGGERPKYCNVCGNRFTTGNILKY 415
Db 308 RLERQGNPTQERRHKKDECGKSFAGSGLVRHNIHTGKERKYQCNVCGKAFYRSALLS 367
QY 416 HHHRRREKPRHVOMNPRVPEHLDVYITSSGLPYMSVPRKEAEBAEPATRGCGVEKKPLV 475
Db 368 HODIHNK-----YKRYNCK 381
QY 476 ASTTLALSTATESLTLTSTGSTAVALRGLTFFNKFFVLMKVAVERPSKADENTPRGSEGSALG 535
Db 382 ECGKAFPSQNTGL-----ILHQRHTGEEKRYQCMOCGKAFFSOSAG 420
QY 536 VADSGSATREMQLSKLVTLSPSWALLTGNLKSSTGSEFPFVYLEPLGASPSTKLOOLEVER 595
Db 421 L-----ILHQRHTGGER--PYECNECGKAFHSHSL----- 449
QY 596 IDRGCAVAVASTASGAPTTAPAPABSSASGPNOCYICLRVLSCPRALRLHYGONGGERPF 655
Db 450 IGHORI-----HTGEEKPYEDEDCCGKTFRRSHILIGNRSHTGEEKRY 490
QY 656 KCKVGRASSTRENTLRAHFVGHKTPSPARAQNSQICQKKFTNATVTLQOHNVHNLG 711
  
```


491 KCNCGRAFSOKSGL-----IEHQRIHTGERPYKCECKGAFNGTGLIQLHRIHTG 542

RESULT 33

2263_HUMAN STANDARD: PRT: 683 AA.

AC 014978; 043387; Created

30-MAY-2000 (Rel. 39, Last sequence update)

15-JUN-2002 (Rel. 41, Last annotation update)

zinc finger protein 263 (Zinc finger protein FPM315).

ZNF263 OR FPM315.

Homo sapiens (Human).

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

NCBI_TaxID=9606;

SEQUENCE FROM N.A.

MEDLINE=97398134; PubMed=9256059;

Yokoyama M., Nakamura M., Okudo K., Matsubara K., Nishi Y., Matsumoto T., Fukushima A.;

Isolation of a cDNA encoding a widely expressed novel zinc finger protein with the Lef and KRAB-A domains.

Biochim. Biophys. Acta 1353:13-17(1997).

(2)

SEQUENCE FROM N.A.

Ricke D.O., Bruce D., Mundt M., Doggett N., Munk C., Saunders E., Robinson D., Jones M., Buckingham J., Chasteen L., Thompson S., Goodwin L., Bryant J., Tesmer J., Melnick L., Longmire J., White S., Ueng S., Tatum O., Campbell C., Fawcett J., Maltbie M., Misra M., Deaven L.;

Submitted (FEb-1998) to the EMBL/GenBank/DBJ databases.

- FUNCTION: MIGHT PLAY AN IMPORTANT ROLE IN BASIC CELLULAR PROCESSES AS A TRANSCRIPTIONAL REPRESSOR.

- SUBCELLULAR LOCATION: Nuclear (By similarity).

- TISSUE SPECIFICITY: EXPRESSED IN HEART, BRAIN, PLACENTA, LUNG, LIVER, SKELETAL MUSCLE, KIDNEY, PANCREAS, SPLEEN, THYMUS, PROSTATE, TESTIS, OVARY, SMALL INTESTINE, COLON AND LEUKOCYTE.

- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-FINGER PROTEINS.

- SIMILARITY: CONTAINS 1 SCAN BOX.

- SIMILARITY: CONTAINS 1 KRAB DOMAIN.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL: D88827; BAA21853.1; -

EMBL: AC004232; AAC24490.1; -

HSSP: P08045; 12NF

Gene: HGNC:13056; ZNF263.

MIM: 604191; -

InterPro: IPR001909; KRAB.

InterPro: IPR003309; Treg-SCAN.

InterPro: IPR00822; ZnfC2H2.

pfam: PF00096; zf-C2H2; 9.

pfam: PF01352; KRAB; 1.

pfam: PF02023; SCAN; 1.

PRINTS: PR00048; ZINCFINGER.

ProDom: PD000003; Znf_C2H2; 9.

SMART: SM00349; KRAB; 1.

SMART: SM00431; LER; 1.

SMART: SM00355; ZNF_C2H2; 9.

PROSITE: PS50805; KRAB; 1.

PROSITE: PS50804; SCAN_BOX; 1.

PROSITE: PS00028; ZINC_FINGER_C2H2_1; 9.

PROSITE: PS50157; ZINC_FINGER_C2H2_2; 9.

Transcription regulation; zinc-finger; Metal-binding; Nuclear protein; DNA-binding; Repeat; Repressor.

FT DOMAIN 41 123 SCAN BOX.

FT DOMAIN 217 289 KRAB.

FT ZINC_FING 378 681 ZINC_FINGERS.

FT ZN_FING 378 400 C2H2-TYPE.

FT ZN_FING 434 456 C2H2-TYPE.

FT ZN_FING 462 484 C2H2-TYPE.

FT ZN_FING 490 512 C2H2-TYPE.

FT ZN_FING 518 540 C2H2-TYPE.

FT ZN_FING 575 597 C2H2-TYPE.

FT ZN_FING 603 625 C2H2-TYPE.

FT ZN_FING 631 653 C2H2-TYPE.

FT ZN_FING 659 681 C2H2-TYPE.

FT CONFLICT 118

SEQUENCE 683 AA; 77240 MW; 2DBFE6C1FCE6793A CRC64;

Query Match 5.0%; Score 267; DB 1; Length 683;

Best Local Similarity 21.0%; Pred. No. 4.3e-05;

Matches 145; Conservative 67; Mismatches 226; Indels 254; Gaps 29;

QY 142 SPKLGATPLPPESTAP-----PPPPPP-----PPPG-VGSGHLNI 177

DB 153 SPSEKLEPETERSPGRLOELLGSPDRDPAQKRALAPWLSPFPEGMEDKEMTG 212

QY 178 PLILEELR-----VLQQRQIQHQMTEQICROYLL-----LGSIGQTVGAPASPS-ELPG 226

DB 213 POLPESLEDVAMKISQIEEMGHODPSKRALSRDVOESTYENDLSLSH-----PSQEVVG 267

QY 227 TGAASSTPLPLPSPDKPAQGTGTTAASSSSSSSSGAEPKPAFFHL-----YH 276

DB 268 TVVGCGGK-----LMDP-----SVOSCKEGLSPRAPPEKEFENLEGVPSVSENTH 315

QY 277 P-----LGSHPSPSVGVGVRSHKPT-----ASPRALPGSDQLASHPLAFPTTGLLA 326

DB 316 POUVLDPDARGVPS-PELGRPHDRSQGDWAPP-EGGEOALA-----ISSLE 358

QY 327 AQCIGAANGLEMAASPGLEPK-----NSGELGYGEV-----ISSLE 364

DB 359 ---GASSGRE-LGRKELQPKKHLCLPGCKNFSNNILRHQRIHAERLCMGVDCE 413

QY 365 KPGR-----HKRCFAKVFYSDSALOILHRSHTGERPKCNVCGNRTT 409

DB 414 IFGGNPRPLSLRAHLGEBRANKLECGKCFSONTHLRHQRHTHGEKPYOCNICGKCFSC 473

QY 410 RGNLKVHHRHREKYPHYQMNPRVPEHLDYVTSSGLPVGMSVPRPEKAEEBAGPRGGV 469

DB 474 NSNL-----HHQRIHTGEKPYKCECKGAFNGTGLIQLHRIHTG-----HT 513

QY 470 ERKPLVASTTALSTATESLTLLSTGTSTAVAPGLTFNKFLVLMKAVERPKSKADENTPGE 529

DB 514 GERPKCECGKSFSSSHLV-----IHERTHERELYPSE 550

QY 530 -GSAIAGVADSGSATRMQSLVTLSPSMALLLNHLKSTGSPFPYVLEPLGASSEFTSK 588

DB 551 CGEAVS-----DSTPPLTNH-----GAHRAE--- 571

QY 589 LQQLVEKIDROGAVAVASTASGAPTASAPAPSSASGPNOCVLCRLVSLCPRALHLVQ 648

DB 572 -KKLFE-----CLTGKSPRSGMHLRHRORT 596

QY 649 HGERPFKCKVCGRAFSIRGNLRANFVGHNTSPARAQNSCPICQKKTNAVTLQOHVAM 708

DB 597 HTGERPKCTLCGENFSRNSNL-----IRHQRIHTGEKPYKCECKGAFNGTGLIQLHRIHT 652

QY 709 HLGGOIPMGGSALSEGGAQERSSSQSTASG 740

DB 653 HT-GERPKCECGESFSSSRSLMSHQRHTG 683

RESULT 34

KRUB_DROME STANDARD: PRT: 845 AA.

ID KRUB_DROME

AC P08155; Q9VHM5;

01-AUG-1988 (Rel. 08, Created)

16-OCT-2001 (Rel. 40, last sequence update)
15-JUN-2002 (Rel. 41, last annotation update)
Kruppel homologous protein 1.
KR-H1 OR KR-H OR CG9167.
Drosophila melanogaster (Fruit fly).
Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
NCBI_TaxID=7227;
[1]
SEQUENCE FROM N.A., DEVELOPMENTAL STAGE, AND ALTERNATIVE SPLICING.
STRAIN=Oregon-R; TISSUE=Embryo, Larva, and Pupa;
MEDLINE=20237577; PubMed=10772791;
"Kruppel homolog, a stage-specific modulator of the prepupal ecdysone response, is essential for Drosophila metamorphosis."
Dev. Biol. 221:53-67(2000).
[2]
SEQUENCE FROM N.A. (ALPHA AND BETA ISOFORMS).
STRAIN=Berkley;
MEDLINE=20196006; PubMed=10731132;
Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
Brandon R.C., Rogers Y.-H.C., Blazee R.C., Champe M., Pfeiffer B.D.,
Wan K.H., Doyle C., Baxter E.G., Heit G., Nelson C.R., Miklos G.L.G.,
Ahl J.F., Aydayani A., An H.-J., Andrews-Pfankuch C., Baldwin D.,
Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
Borkov D., Botchan M.R., Butler H., Cadieu E., Center A., Chandra I.,
Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
De Pablos R., Delcher A., Deng Z., Mays A.D., Dew I., Diez S.M.,
Dodson K., Dou P.L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
Glocke A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
Houston D., Houston K.A., Howland T.J., Wei M.-H., Idegawa C.,
Jatani M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Kelchum K.A.,
Klimel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
Liu X., Mettel B., McIntosh T.C., McLeod M.P., McPherson D.,
Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
Palazzo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
Svrtkars R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
Wang Z.-Y., Wasserman D.A., Welnslock G.M., Weissbach J.,
Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
Zheng X.H., Zhong F.N., Zhou W., Zhou X., Zhu X., Smith H.O.,
Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
"The genome sequence of Drosophila melanogaster."
Science 287:2185-2195(2000).
[3]
SEQUENCE OF 289-367 FROM N.A.
MEDLINE=87051757; PubMed=3096579;
Schuh R., Atcher W., Gaul U., Cote S., Preiss A., Maier D.,
Seifert E., Nauber U., Schroeder C., Kemler R., Jaekle H.;
"A conserved family of nuclear proteins containing structural elements of the finger protein encoded by Kruppel, a Drosophila segmentation gene."
Cell 47:1025-1032(1986).
[4]
-1- FUNCTION: PLAYS A GENERAL ROLE IN THE HIERARCHIES OF GENE
EXPRESSION LEADING TO METAMORPHOSIS.
-1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; ALPHA AND BETA (SHOWN HERE); ARE
PRODUCED BY ALTERNATIVE SPLICING.
-1- DEVELOPMENTAL STAGE: BETA ISOFORM IS EXPRESSED DURING

CC EMBRYOGENESIS, MOST ABUNDANT IN MIDEMBRYOGENESIS, AND IN ADULTS.
CC ALPHA ISOFORM IS EXPRESSED FROM EMBRYOGENESIS TO 8 HOURS AFTER
CC PUPARIATION. MAJOR PERIOD OF EXPRESSION IS DURING INSTAR.
CC SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-
CC FINGER PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>
CC or send an email to license@sdb.ch).
CC -----
CC EMBL: AJ005440; CAA06543.2; -
CC EMBL: AJ005441; CAA06544.2; -
CC EMBL: AE003612; AAF52343.2; -
CC EMBL: AE003612; AAG22417.1; -
CC EMBL: M14940; AAA28660.1; -
CC HSP: P08047; 1SP2.
CC P1yBase: FBbrn0028420; Kr-h1.
CC InterPro: IPR000822; Znf-C2H2.
CC Pfam: PF00096; Zf-C2H2; 8.
CC ProDom: PD000003; Znf-C2H2; 4.
CC SMART: SM00355; Znf-C2H2; 8.
CC PROSITE: PS00028; ZINC_FINGER_C2H2_1; 8.
CC PROSITE: PS00157; ZINC_FINGER_C2H2_2; 8.
CC ZINC-finger; Metal-binding; DNA-binding; Repeat; Alternative splicing;
CC developmental protein.
CC FT DOMAIN 108 148 GLN-RICH.
FT DOMAIN 482 515 SER/THR-RICH.
FT DOMAIN 538 565 SER/THR-RICH.
FT DOMAIN 194 463 ZINC-FINGERS.
FT ZN_FING 194 216 C2H2-TYPE.
FT ZN_FING 271 293 C2H2-TYPE.
FT ZN_FING 299 321 C2H2-TYPE.
FT ZN_FING 327 349 C2H2-TYPE.
FT ZN_FING 355 377 C2H2-TYPE.
FT ZN_FING 383 407 C2H2-TYPE.
FT ZN_FING 413 435 C2H2-TYPE.
FT ZN_FING 441 463 C2H2-TYPE.
FT VARSPLIC 1 54 MISSING (IN ISOFORM ALPHA).
FT CONFLICT 313 313 N -> D (IN REF. 3).
SQ SEQUENCE 845 AA; 91451 MW; A4d878E98DC8E372 CRC64;
Query Match 5.0%; Score 266; DB 1; Length 845;
Best Local Similarity 18.2%; Pred. No. 5.9e-05;
Matches 173; Conservative 87; Mismatches 328; Indels 362; Gaps 31;
QY 163 PPP-----PPGVSGHLN-----PLIEELRVIQ 188
DB 80 PPLPLATTTVSGLVGVPSSGQGHFELLQTPQQRQMLQLODHOQEQGFVYQALIQ 139
QY 189 QROTHOMQTEOICROVLLISLGQIV-----GAPASPELPGGASTRKLLPLFSP 242
DB 140 HQKQOQOQOQHESITNAPTAAPSAQRKITEPVGGFPAASAAVVSQYKRPASRP-----Q 193
QY 243 IKPAQGTKTASSSSSSSSSGAEPKRAF-FLYHPLISQHPFSVGVRSHKPTAPSP 301
DB 194 FKQDQCGMTFGSKSARTSHTKSHKNODLSNGASGAVAPVSTRAILINDAGLPGIP 253
QY 302 ALPGSTDLIASPHLAFPGTTGLLAQCLGARGLEAASAPGLKPKNGSGELGYEVIS 361
DB 254 KSP--TKRLAV-----AAGADPYC-INVQQTFAVPARLIRHYRRTGGERP----- 299
QY 362 SLEKPGRHKRCFAVFGSDALQILHRSITGTERPYKCNVCGNFTTGNLKVHHRHR 421
DB 300 -----ECECHKLFYKENVQVHRHRTKRPYKCDVCGRAFEHSGKLRHMRHT 350
QY 422 EKYPVQNMHPVPEHLVDVYITSSGLPYGMSVPEKAEAEACTPGGVRKPLVASTAL 481
DB 351 GERPH----- 355

482 SATESLTLSTGTSTAVAPGLPTFNKFKAVEPKSKADENTPPGSESAIAGVADSGS 541
356 -----KQSVCKT-----FIQSGQ 369
542 ATRMQLSKLVTLSPSWALLTLNLKS-TGSEPPPYVLEPLCASPSETSKLQOLVERKIDRG 600
370 -----LVIMHRTHTGKPY----- 383
601 AVAVASTASGAPPTSPAPASSSSASGPNQVICLRVLSCPRALRLHYOGHGERPKKVC 660
384 -----KCPERG-----CGKGFYCSQKLVKSHRTHTGKPYHCDIC 418
661 GRASTRGRLNAHFVGHKTSAPARAQNSCPICOKKFTNAVTLQOHVRMLGQIPN---- 716
419 FRDGYNVHVLKRVQHGSKCYK---CTICDETFFNKKKMEAHIGH-ANEVDDENE 473
717 -----GGALSEG--GGAAGENSSEOSTASGPGSPPOP-OSQP----- 752
474 AAAAAASTASGSSAGSAGSLQGVSSNSNSHSPSPATKPRARQPRVSKTVAT 533
753 -----SPE-EMSEEEDEEEDVDDESLAGSGSGEKATSV 793
534 LSTPSSPLSSLSSTYSPSASSMAAPPPTSAHYLPVQMEADALSRDGVSSAQAHPST 593
794 RGDSEVSGAESEVATSVAAPTTVKEMDSNEKARPHTLPPPPPPDNL-HPOPMEOGTS 852
594 YADBE-----PTDLSMQVQGLPESTVDYVQAPPSLLELQPP----- 632
853 DVSQAMEEAKLEGISSPMALTOGEGSTPLVEELNLEAMKKDEGSGKACVCG 912
633 ---AGLTINPALLEA---ASTARRHDDDDQVODEVDVHAAMQM-----MOLCRGHG 678
913 QSEPTQALIEHQ---KTHPKDGLFTFCVFCRQGLDRATLKHKMLLHNOVPFAPHG 968
679 SLRPTQRPAPSHQPVPTLHVSD-----LAANDD--THEATVILEHKKRGLAHNG 728
969 -----PONTATLSLVPCSSSIPSPGSLPPFRKDDPTM 1001
729 LHKGAPVPKYESALPNPDVVARVEAIGLRSTESPERSSPESDSIMM 778
RESULT 35
ZNF43_HUMAN STANDARD; PRT; 803 AA.
ID ZNF43_HUMAN STANDARD; PRT; 803 AA.
AC P17038; P28160; Q96DG1;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Zinc finger protein 43 (Zinc protein Htf6) (Zinc finger protein
Kox27).
EN ZNF43 OR ZNF39 OR KOX27.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-T-cell;
RA MEDLINE=91279444; PubMed=1711675;
RT "A gene encoding 22 highly related zinc fingers is expressed in
lymphoid cell lines."
RL Nucleic Acids Res. 19:2921-2927(1991).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE-Lymph;
RA Strausberg R.;
RN Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 38-190 FROM N.A.
RC MEDLINE=91219421; PubMed=2023909;
RA Bellefroid E.J., Poncetlet D.A., Lecocq P.J., Revelant O.,
RA Martial J.A.;

RT "The evolutionarily conserved Kruppel-associated box domain defines a
RT subfamily of eukaryotic multifingered proteins.";
RL Proc. Natl. Acad. Sci. U.S.A. 88:3608-3612(1991).
RN [4]
RP SEQUENCE OF 476-531 FROM N.A.
RC TISSUE-Lymphoid;
RX MEDLINE=91145339; PubMed=2288909;
RA Thiesen H.-J.;
RT "Multiple genes encoding zinc finger domains are expressed in human T
RT cells.";
RL New Biol. 2:363-374(1990).
CC -1- FUNCTION: MAY BE INVOLVED IN TRANSCRIPTIONAL REGULATION.
CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -1- TISSUE SPECIFICITY: T AND B CELL LINES.
CC -1- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-
CC FINGER PROTEINS.
CC -1- SIMILARITY: CONTAINS 1 KRAB DOMAIN.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X59344; CAA41932.1; -
CC EMBL; BC006528; AAH06528.1; -
CC EMBL; M61869; AAB58674.1; -
CC EMBL; X52358; CAA36584.1; -
CC PIR; S26823; S26823.
CC PIR; D39384; D39384.
CC PIR; S10416; S10416.
CC HSP; P08048; 72NMF.
CC TRANSFAC; T04986; -
CC Genew; HGNC; 13109; ZNF43.
CC MIM; 603972; -
CC InterPro; IPR001909; KRAB.
CC InterPro; IPR000822; Znf_C2H2.
CC Pfam; PF00096; Zf-C2H2; 21.
CC Pfam; PF01352; KRAB; 1.
CC PRINTS; PR00048; ZINCFINGER.
CC PRODOM; PD000003; Znf_C2H2; 16.
CC SMART; SM00349; KRAB; 1.
CC SMART; SM00355; Znf_C2H2; 21.
CC PROSITE; PS50805; KRAB; 1.
CC PROSITE; PS00028; ZINC_FINGER_C2H2_1; 19.
CC PROSITE; PS50157; ZINC_FINGER_C2H2_2; 22.
KW Transcription regulation; DNA-binding; Zinc-finger; Metal-binding;
KW Nuclear protein; Repeat.
FT DOMAIN 1 69 KRAB.
FT ZN_DOMAIN 168 778 ZINC_FINGERS.
FT ZN_FING 168 190 C2H2-TYPE.
FT ZN_FING 196 218 C2H2-TYPE (DEGENERATE).
FT ZN_FING 224 246 C2H2-TYPE (DEGENERATE).
FT ZN_FING 252 274 C2H2-TYPE (DEGENERATE).
FT ZN_FING 280 302 C2H2-TYPE.
FT ZN_FING 308 330 C2H2-TYPE.
FT ZN_FING 336 358 C2H2-TYPE.
FT ZN_FING 364 386 C2H2-TYPE.
FT ZN_FING 392 414 C2H2-TYPE.
FT ZN_FING 420 442 C2H2-TYPE.
FT ZN_FING 448 470 C2H2-TYPE.
FT ZN_FING 476 498 C2H2-TYPE.
FT ZN_FING 504 526 C2H2-TYPE.
FT ZN_FING 532 554 C2H2-TYPE.
FT ZN_FING 560 582 C2H2-TYPE.
FT ZN_FING 588 610 C2H2-TYPE.
FT ZN_FING 616 638 C2H2-TYPE.
FT ZN_FING 644 666 C2H2-TYPE.
FT ZN_FING 672 694 C2H2-TYPE.
FT ZN_FING 700 722 C2H2-TYPE.
FT ZN_FING 728 750 C2H2-TYPE.

Best Local Similarity 23.9% Pred No. 6,1e-05;
Matches 96; Conservative 42; Mismatches 141; Indels 123; Gaps 13

351 SGEIAGEVYISLSEKDGRIHRCFAKAVFSGDSALQILNRSHTGERPYKCVCNGNRFTTR 410
| : : : | : : | : : | : : | : : | : : | : : | : :
290 SKSLIOHQRIHTGEEKP---YRCECGKAFCOSSSLIHQRHTHGEPYGCRCCKAFSSQ 346
411 GNLUKVHFHHRRKKXPYQNMPHPREHLDYITTSGLRYGNSVPPEKAEDEAGTGGGVE 470
| : : : | : : | : : | : : | : : | : : | : : | : :
347 SOLV-----RHRTTHTEGPYPCKE-----CGKAFS 372
471 RKLVAFTALSTESTLLSTGTSTAVAPGLPTFNKVLKMAVEPKSKADENTPPSGEG 530
| : : : | : : | : : | : : | : : | : : | : : | : :
373 QSSTLAQHQMHNHGKAKQLIKASDPSLVAHQR---IHAWEKPRCKDEC-----G 419
531 SATAGVADSASATRMOL-----SKLYTSLPSMALLNHLKS-TGSFPF----- 572
420 KAFRWIS--RLSOHOLIHTGEEKPYCNKCKPAFGCCSSRLRNHQTHTGEEKPFKDECGK 476
573 -----PYLEBPLGASPSETSKL--QQLVERIDROGAIVASTASGA 611
477 GFPOGSHLIHQRIHTGEEKRPVNCDCCKAFCOSSSLIHQRHKKEK----- 523
612 PPTSAPASSASGNOCVICLRVLSCPRALRLHYGHGGERPFCKVKCYGRAFSSTRGNLR 671
524 -----PYELCCKGKAFMSSTOLTIIHQRVHTGEPYKCNECGKAFSONSTLF 569
672 AHEVGKHTSPARAQNOSCPIOCKKFTNAVTLOQHVRLMLGGQ 713
510 QHQIITH---AGVKPYECSECKGAFMSNSTYLIEHORHTTRAQ 607

RESULT 37
ID 2151 HUMAN STANDARD; PRT: 803 AA.
15-JUL-1998 (Rel. 36, Created)
15-JUL-1998 (Rel. 36, Last sequence update)
15-JUN-2002 (Rel. 41, Last annotation update)
zinc finger protein 151 (Myc-interacting zinc finger protein) (Miz-1 protein).
ZNF151 OR MIZ1.
Homo sapiens (human).
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
NCBI_TaxId=9606;
[1]
SEQUENCE FROM N.A.
MEDLINE=97453565; PubMed=9308237.
Schneider A., Peukert K., Ellers M., Haenel F.:
"Association of Mgc with the zinc-finger protein Miz-1 defines a novel pathway for gene regulation by Mgc.";
novel pathway of Mgc regulation by Mgc.";
Curr. Top. Microbiol. Immunol. 224:137-146(1997).
[2]
SEQUENCE FROM N.A.
Bird C.;
Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
[3]
SEQUENCE OF 580-803 FROM N.A.
TISSUE=Insulinoma;
MEDLINE=96044430; PubMed=7557990;
Tommerup N., Vissing H.;
"Isolation and fine mapping of 16 novel human zinc finger-encoding cDNAs identify putative candidate genes for developmental and malignant disorders.";
Genomics 27:259-264(1995).
[4]
SEQUENCE OF 327-342 FROM N.A.
TISSUE=Placenta;
MEDLINE=92372070; PubMed=1505991;
Lichter P., Bray P., Ried T., David I.B., Ward D.C.;
"Clustering of C2-H2 zinc finger motif sequences within telomeric and fragile site regions of human chromosomes.";

DR	Genomics13:999-1007(1992).
CC	-1- SUBUNIT: ASSOCIATED WITH THE C-TERMINAL OF MYC.
CC	-1- SUBCELLULAR LOCATION: Nuclear (potential).
CC	-1- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-FINGER PROTEINS.
CC	-1- SIMILARITY: CONTAINS 1 BTB/POZ DOMAIN.
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC	EMBL; U20647; AAC50256.1; -
DR	EMBL; AL034555; CAB85445.1; -
DR	EMBL; Y09723; CAI70889.1; -
DR	EMBL; M88369; AAA61327.1; -
DR	HSSP; P07248; IARE.
DR	TRANSFAC; T03414; -
DR	GeneW; HGNC:12936; ZNF151.
DR	MIM; 604084; -
DR	InterPro; IPRO00210; BTB_POZ.
DR	Interpro; IPR000822; ZnF_C2H2.
DR	Pfam; PF00096; zf-C2H2; 13.
DR	Pfam; PFO0651; BTB; 1.
DR	PRINTS; PR00048; ZINCFINGER.
DR	ProDom; PD000003; ZnF_C2H2; 2.
DR	SMART; SM00225; BTB; 1.
DR	SMART; SM00355; ZnF_C2H2; 13.
DR	PROSITE; PS50097; BTB; 1.
DR	PROSITE; PS00028; ZINC_FINGER_C2H2_1; 13.
DR	PROSITE; PS50157; ZINC_FINGER_C2H2_2; 13.
KW	Transcription regulation; DNA-binding; Zinc-finger; Metal-binding;
KW	Nuclear protein; Repeat.
FT	DOMAIN 1 104 BTB.
FT	DOMAIN 306 739 ZINC_FINGERS.
FT	ZN_FING 306 739 C2H2-TYPE.
FT	ZN_FING 334 356 C2H2-TYPE.
FT	ZN_FING 362 384 C2H2-TYPE.
FT	ZN_FING 390 412 C2H2-TYPE.
FT	ZN_FING 418 440 C2H2-TYPE.
FT	ZN_FING 446 468 C2H2-TYPE.
FT	ZN_FING 474 496 C2H2-TYPE.
FT	ZN_FING 502 524 C2H2-TYPE.
FT	ZN_FING 530 552 C2H2-TYPE.
FT	ZN_FING 558 580 C2H2-TYPE.
FT	ZN_FING 586 608 C2H2-TYPE.
FT	ZN_FING 614 637 C2H2-TYPE.
FT	ZN_FING 717 739 C2H2-TYPE.
FT	CONFLICT 73 M->V (IN REF. 2).
SEQ	SEQUENCE 803 AA; 87959 MW; FEE72BA4E14AE200 CRC64;
QY	Query Match 5.0%; Score 263; DB 1; Length 803;
QY	Best Local Similarity 19.3%; Pred. No. 7.5e-05;
QY	Matches 178; Conservative 92; Mismatches 312; Indels 342; Gaps 31.
DG	73 SSASSAPRPECHSR-----OVMDEHSNPDSGSGSPDPDWGPERRKEE 118
DG	113 SLAEPATSPGCNAEALATEGGCKRAKEEVAVSTLSRLDAQGRSTP---IGPSIDLKEE 168
DG	119 SSGGFVAATGTAAAGGGGLLASPKLATPTLPPESTPAAPPDPPPP-----PPPPVGSGH 174
DG	169 RGGDAQSASGAE-----QTEKADAPREPPVELKPDPTSGMAAAE 209
DG	175 LNIPLI-----LEELRVLAQRDIHQOMTEIQCVLLIGSLGQTGVGARPSSELR 226
DG	210 AEALESSESQEWEVEPARKEGEEOKEOEEOEEE-----GAGPAEYKE 252
DG	227 TGAASSTRPLRLPFPIKPAOTGKTITASSSSSSSSGAAPPKAFPHLHPDGSHPSV 286
DG	253 EG-----SOLENGAPEENENEESACTDSGE-----LGSE----- 283


```

287 GGVGRSHKTPAPSPALPGSTDLIASPHLAFPGTGTLLAQCCLAGRLAEMASPGLK 346
284 -----AKGLRSSTGYDRT 297
347 PRKNGSGELGYGEVVISLEKPRGGRHKRCFAKVFSGSDALQIHLRSHTEGRPYCNVGNR 406
298 SK-----AYGCVI-----HKCEDCGKEFTHTGNFKRIIRIHTGKPKFCRCRCSKA 342
407 FTTGNGNLKVHFRHREKYPHVQNMNPHVPEHLDVITSSGLPYGMSVPEKAEKSTPG 466
343 FSDPACAAHEKTH-----SPLKPYGC-----EECG-- 368
467 GGVGRKPLVASTLALSAFESLTLSTGTSTAVAGLPPTNKFVLMKAVEPKSKADENTPP 526
369 -----KSYRLIS-----LNLHK-----K 382
527 GSEGSALAGVADSGSATRNQSLKVLTLSPFWALLTNHLKSTGSPFPVYLEPLGASPE- 585
383 RHSEGARVRCEDCG-----KLFTT--SGNLKRHQLVHSGE--KPYCCDYCGRGSFSDP 430
586 TSKRLOOL-VERIDRGAVAVASTASGAPTSAPAPSSASGPNOCVILCLRVLSCPALRL 644
431 TSKRHLTETHDTKE-----HKCPHCDKKNQVGNLKA 463
645 HYGONGGERPRKCKVCGRAFSTGRLRAHFVGHKTSAPARAQNSCPICOKKFTNAVTLQ 704
464 HLKTHIDGPLKRCRCGKQFTSGNLKRHLRIH-----SGEKPYVICHORORADPALOR 519
705 HYRHLHG---QIPNGSALSCEGGAOENSSSEOSTASGSPSPQPOSOQSPDEEMSEE 761
520 HVRHTEKRPQCVNCGKAFTQ-----ASSLIHVRHTEKRPVYCERCGRK 566
762 EEEDEEEEDVTDSDSLAGRSGESGGEKATSVRGDSEVSAAEEVATSVAPTYKEMD 821
567 FVQSSQLANLIRHNHNIRH-----KCSVCSKAFNVGD 600
822 SNEKAPQHTLPPR-----PPDNT-DHPQPMQGTSDVSGAMEEAKLEGISSPM 871
601 LSKHIIHTGKRPYLCDCGRGFRNVDNLRHVKTVHKGKAGIKLEPEEGSESVVYTD 660
872 AALQEGEGSTPLVEELN-LPEAMKKDRESSGRKACEVCGQSPPTALPEHQKTHK 930
661 DMTLALAEALATATVTLVVGAAVTADETVLKA-----EISKAVQVQVEDPN 712
931 DGPLFTCVFCRQGLDRLATLKKHM 954
713 THILYACDSCGDKFLDANSIAGHV 736
SUPL 38
008_HUMAN STANDARD: PRT; 1167 AA.
043345;
16-OCT-2001 (Rel. 40, Created)
15-OCT-2001 (Rel. 40, Last sequence update)
15-JUN-2002 (Rel. 41, Last annotation update)
Zinc finger protein 208.
ZNF208 OR ZNF91L.
Homo sapiens (Human).
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheta; Primates; Catarrhini; Hominiidae; Homo.
NCBI_Taxid=9606;
[1]
SEQUENCE FROM N.A.
MEDLINE=98391764; Pubmed=9724325;
Eichler E.F., Hoffman S.M., Adamson A.A., Gordon L.A., McCreedy P.,
Lamerdin J.E., Mohrenweiser H.W.,
"Complex beta-satellite repeat structures and the expansion of the
zinc finger gene cluster in 19p12.";
Genome Res. 8:791-808(1998).
-!- FUNCTION: MAY BE INVOLVED IN TRANSCRIPTIONAL REGULATION.
-!- SUBCELLULAR LOCATION: Nuclear (Probable).

```

```

CC -!- TISSUE SPECIFICITY: UBIQUITOUS.
CC -!- SIMILARITY: BELONGS TO THE KRUPPEL FAMILY OF C2H2-TYPE ZINC-
CC FINGER PROTEINS.
CC -!- SIMILARITY: CONTAINS 1 KRAB DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AC003973; AAB94784.1; -.
CC HSSP: P25490; 1UBD.
CC Genew: HGNC:12999; ZNF208.
CC MIM: 606760; -.
CC InterPro: IPR001909; KRAB.
CC InterPro: IPR000822; ZnF_C2H2.
CC Pfam: PF00096; zf-C2H2; 33.
CC Pfam: PF01352; KRAB; 1.
CC PRINTS: PR00048; ZINC_FINGER.
CC ProDom: PD000003; ZnF_C2H2; 17.
CC SMART: SM00349; KRAB; 1.
CC SMART: SM00355; ZnF_C2H2; 33.
CC PROSITE: PS50805; KRAB; 1.
CC PROSITE: PS50028; ZINC_FINGER_C2H2_1; 33.
CC PROSITE: PS50157; ZINC_FINGER_C2H2_2; 34.
CC KW Transcription regulation; DNA-binding; Zinc-finger; Metal-binding;
CC Nuclear protein; Repeat.
CC FT DOMAIN 7 75
CC FT ZN_FING 201 1147 KRAB.
CC FT ZN_FING 201 223 ZINC_FINGERS.
CC FT ZN_FING 229 251 C2H2-TYPE (DEGENERATE).
CC FT ZN_FING 257 279 C2H2-TYPE.
CC FT ZN_FING 285 307 C2H2-TYPE.
CC FT ZN_FING 313 335 C2H2-TYPE.
CC FT ZN_FING 341 363 C2H2-TYPE.
CC FT ZN_FING 369 391 C2H2-TYPE.
CC FT ZN_FING 397 419 C2H2-TYPE.
CC FT ZN_FING 425 447 C2H2-TYPE.
CC FT ZN_FING 453 475 C2H2-TYPE.
CC FT ZN_FING 481 503 C2H2-TYPE.
CC FT ZN_FING 509 531 C2H2-TYPE.
CC FT ZN_FING 537 559 C2H2-TYPE.
CC FT ZN_FING 565 587 C2H2-TYPE.
CC FT ZN_FING 593 615 C2H2-TYPE.
CC FT ZN_FING 621 643 C2H2-TYPE.
CC FT ZN_FING 649 671 C2H2-TYPE.
CC FT ZN_FING 677 699 C2H2-TYPE.
CC FT ZN_FING 705 727 C2H2-TYPE.
CC FT ZN_FING 733 755 C2H2-TYPE.
CC FT ZN_FING 761 783 C2H2-TYPE.
CC FT ZN_FING 789 811 C2H2-TYPE.
CC FT ZN_FING 817 839 C2H2-TYPE.
CC FT ZN_FING 845 867 C2H2-TYPE.
CC FT ZN_FING 873 895 C2H2-TYPE.
CC FT ZN_FING 901 923 C2H2-TYPE.
CC FT ZN_FING 929 951 C2H2-TYPE.
CC FT ZN_FING 957 979 C2H2-TYPE.
CC FT ZN_FING 985 1007 C2H2-TYPE.
CC FT ZN_FING 1013 1035 C2H2-TYPE.
CC FT ZN_FING 1041 1063 C2H2-TYPE.
CC FT ZN_FING 1069 1091 C2H2-TYPE.
CC FT ZN_FING 1097 1119 C2H2-TYPE.
CC FT ZN_FING 1125 1147 C2H2-TYPE.
CC SQ SEQUENCE 1167 AA; 134352 MW; E2184DF23B0D35E9 CRC64;
Query Match 5.0%; Score 263; DB 1; Length 1167;
Best Local Similarity 21.3%; Pred. No. 0.00011;
Matches 147; Conservative 84; Mismatches 284; Indels 176; Gaps 26;
Oy 358 EVISLEKPGRRHKRCFAKVFSGSDALQIHLRSHTEGRPYCNVGNRPTTRGNLKVHF 417

```


416 EYIHTGKRP---YKCECGKAFNWSNLMENKIHGTGPRYKCECGKGFMSSTLSYHK 472
418 HHRREKYPRHQMNDPVRPHLDVYTSSGLPGMSV-----PREKAEFGTP 465
473 KIHTEVKP-----YKCECGKAFNWSAILIKHKRIHTEVKPKEE---C 514
466 GGGVERKPLVASTTALSATE-----SLPTLSTGTST---AVAPGL-----PTF 505
515 GKTFSKYSTLTTHKAIHAGEKPYKCEGKFTIKYSTLTTHKAIHAGEKPYKCEGKAF 574
506 NKFVLM---KAV---EPKSKADENTPPGSEGSALAGVADSGSATRMQLSKLYT----- 552
575 SKFSILTTHKHVIHTEGKPYKCEE-----CGKAFNWSNLMENKIHTEGKPYKC 623
553 -----SLPSWALLTNHLSKSTGSPFPYVLEPLGASPSSETSLO-----QVLEKIDRGAV 602
624 EECGKSFSTFSVLTKH-KVHTEGKPYKCECGKAYKMSSTLSYHKIHTEVKPKEEC 682
603 AVASTASGAPTTAPAPSSASGPNOCVLCRLVLSCPRALRLHYGONGEPREPKCYGGR 662
683 GKAFNSALILIKHKRIHTEDEK---PYKCECGKTFSKVSTLTTHKAIHAGEKPYKCEGK 740
663 ASTRONLNAHFHGHKTSAPARAQNSCPTQCKFTNAVTLQOHVHMLGGQIDNGSALS 722
741 AFSKTSILTKHKHVIH---TGKPYKCECGKAYKWPSTLSYHKIHT-GEKPYKCECG 795
723 ECGGAQENSSEOSTASGSPFPQPOSOOPREEMSEEEDEEEDVTDSDSLARG 782
796 KQFMSFILTTHKHVIHTEG-----EKPYKCECG----- 823
783 SESGGEKA---ISVRDSEVSGAE-----EVATSVAAPTTVKEMDSNEKAPDHTL 831
824 -----KAFSMLSTVFSKHKTAGEKPYKCEAGKAYMTFSLTKHKVHTEGK----- 871
832 PPPPPRDLNDHQPMQ---GTSVYSGAMEEAKLEGISSPMAALTQOEGSTPLVEEL 889
872 -----PYKCECGKAFNWSNLMENKIHTEGTPYKC---EECD-----KAF 910
890 NLPKAMKPDGSESSGRK--ACEVCGSEFPTQTALEEHQKTHKDDGLFVCPRCQFLDR 947
911 SMPSSLTEHKATHAGEKPYKCECGKAFMSRLETKATHAGEEP-YKCECGKAFNWS 969
948 ATLKKHMLLAHHOVPAPRAPHONIAATLSLV 978
970 SNLMENKRIHTEGKPYKCECGKSFSTLSIL 1000
RESULT 39
LAS_DROME
D GLAS_DROME STANDARD; PRT; 604 AA.
C P13360;
T 01-JAN-1990 (Rel. 13, Created)
T 01-JAN-1990 (Rel. 13, Last sequence update)
T 01-OCT-1996 (Rel. 34, Last annotation update)
E Glass protein.
N GL.
N Drosophila melanogaster (Fruit fly).
N Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
N Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
N Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
N NCBI_TaxID=7227;
N [1]
P SEQUENCE FROM N. A.
P MEDLINE=89365138; PubMed=2770860;
P Moses K., Ellis M.C., Rubin G.M.;
P "The glass gene encodes a zinc-finger protein required by Drosophila
P photoreceptor cells".
P Nature 340:531-536(1989).
P -!- FUNCTION: GLASS IS PROBABLY A TRANSCRIPTION FACTOR REQUIRED FOR
P GENE EXPRESSION SPECIFIC TO PHOTORECEPTOR CELLS.
P -!- SUBCELLULAR LOCATION: Nuclear (Potential).
P -----
P C -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: X15400; CA33450.1; -.
DR PIR: S05447; S05447.
DR HSSP: P08046; 1A1I.
DR TRANSFAC: T00329; -.
DR Flybase: FBgn0004618; g1.
DR InterPro: IPR000822; znf_C2H2.
DR Pfam: PF00096; zf-C2H2; 5.
DR PRINTS: PR00048; ZINC_FINGER.
DR ProDom: PD000003; znf_C2H2; 2.
DR SMART: SM00355; znf_C2H2; 5.
DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; 5.
DR PROSITE: PS0157; ZINC_FINGER_C2H2_2; 5.
DR Transcription regulation; Zinc-finger; Metal-binding; DNA-binding;
KW Repeat; Vision; Nuclear protein.
KW DOMAIN 437 571 ZINC_FINGERS.
FT ZN_FING 437 459 C2H2-TYPE.
FT ZN_FING 465 487 C2H2-TYPE.
FT ZN_FING 493 515 C2H2-TYPE.
FT ZN_FING 521 543 C2H2-TYPE.
FT ZN_FING 549 571 C2H2-TYPE.
FT VARIANT 60 60 E -> Q.
FT VARIANT 361 362 PM -> SL.
FT VARIANT 377 377 S -> T.
SQ SEQUENCE 604 AA; 62528 MM; 82B947C78366C071 CRC64;

Query Match 4.98; Score 262; DB 1; Length 604;
Best Local Similarity 19.3%; Pred. No. 6.3e-05;
Matches 125; Conservative 48; Mismatches 171; Indels 304; Gaps 18;

QY 118 ESSGQFLVANTGTAAGGGGLIASPKLGATPLPESTAPRPPPPPPPP----- 168
DB 231 QSSGNNIGSNGSA-----GVANMAACYET--SAGTAAPPPPAAMYPMSVNSMN 283
QY 169 -----GVSGHL-NIPILIEELRVLDQRIHQOMQTEQICRQVLLGSLGQVGAAPS 222
DB 284 MTMHHGVGADAGGVPM-----QCSGMNT-----PPSNS 313
QY 223 ELRGTAASATKRLPLPSPIKPAQTGKTATSSSSSSSSSGAEPKQAFNLVHPLGSH 282
DB 314 TSAAMAAAVNVLPRLSP-----GHYPASATYSFTADFRAPARTG-----LGALP 360
QY 283 PFSVGVGSRSHKPPAPSPPL-----PGSTDLIASPHIAFGTTGLLAQCICGAARGLE 337
DB 361 PWT---VGEKESPPRANSSLAGYPTGVGNGGYTPPHKS-----PTSYAALGLS 409
QY 338 AAASPLILKPKNSGELGVVISLEKPRGRHK--CRFCAKVFSDSALQIHLSHTG 394
DB 410 LSA-----FDEEDSNEEDLDQDESS-----GGEMKNILCRLCCKTARPTLKHILRHSG 461
QY 395 ERPKYKNCVGNRTTQGNLKVHPIHREKRYHVQKNPVRPHLDVYTSSGLPYGNSVP 454
DB 462 ERYRYPDCCKSSQAANLTAHVRTH----- 487
QY 455 PEKAEAEAGTPGGGVVERKPLVASTTALSATESLTLSTGTSTAVAVAPGLPTFNKFLMKAV 514
DB 488 ----- 487
QY 515 EPKSKADENTPPGSEGSALAGVADSGSATRMQLSKLYTSLPSWALLTNHLSKSTGSPFPY 574
DB 488 -----TGQKPF--- 493
QY 575 VLEPLGASPSSETSLOLVKIDRGAVAVASTASGAPTTAPAPSSASGPNOCVLYC 634
DB 494 -----RCPICDR 500

635 VLSCPALRLHYGCHGGRPFCKYCGRAFSTRGNLRAHFVGHKTSPPARAONSCPIQOK 694
 501 RFSQSSSVTHMRTSHSGRPYCCSKKSFSDSSTLTKHLRIH-----SGEKPYQCKLCLL 556
 695 KFPNAVTLQOHVHMHGQIIPNGSGASGAQAGENSSEOSTASPG 742
 557 RFSQSGNLNRMHRYHGNNSNGSNATGVG-----ESTSGGVG 597

SULT 40

93 MOUSE
 ZF93 MOUSE STANDARD PRT 645 AA.

061116, 16-OCT-2001 (Rel. 40, created)
 16-OCT-2001 (Rel. 40, last sequence update)
 16-OCT-2001 (Rel. 40, last annotation update)
 Zinc finger protein 93 (Zfp-93).
 ZFP93.
 Mus musculus (Mouse).
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NCBI_TaxID=10090;
 [1]
 SEQUENCE FROM N.A.
 STRAIN=CD-1; TISSUE=Testis;
 MEDLINE=96207307; PubMed=8617494;
 Shannon M., Ashworth L.K., Mucenski M.L., Lamerdin J.E., Branscomb E.,
 Studts L.;
 "Comparative analysis of a conserved zinc finger gene cluster on human
 chromosome 19q and mouse chromosome 7.";
 Genomics 33:112-120(1996).
 -1- FUNCTION: MAY FUNCTION AS A TRANSCRIPTION FACTOR.
 -1- SUBCELLULAR LOCATION: Nuclear (Probable).
 -1- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-
 FINGER PROTEINS.
 -1- SIMILARITY: CONTAINS 1 KRAB DOMAIN.

 This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

EMBL: U46186; AAE03529.1; -
 HSSP: P25490; 1ZNM.
 MGD: MGI:107611; Zfp93.
 InterPro: IPR001909; KRAB.
 InterPro: IPR000822; Znf.C2H2.
 Pfam: PF00096; zfc2h2; 13.
 Pfam: PF01352; KRAB; 1.
 PRINTS: PR00048; ZINC_FINGER.
 PRODOM: PD000003; Znf.C2H2; 11.
 SMART: SM00349; KRAB; 1.
 SMART: SM00355; Znf.C2H2; 13.
 PROSITE: PS00805; KRAB; 1.
 PROSITE: PS00028; ZINC_FINGER_C2H2_1; 13.
 PROSITE: PS50157; ZINC_FINGER_C2H2_2; 13.
 Transcription regulation; DNA-binding; Zinc-finger; Metal-binding;
 Nuclear protein; Repeat.
 DOMAIN 8 86 KRAB.
 ZINC_FINGERS.
 ZN_FING 285 643 C2H2-TYPE.
 ZN_FING 307 643 C2H2-TYPE.
 ZN_FING 313 335 C2H2-TYPE.
 ZN_FING 341 363 C2H2-TYPE.
 ZN_FING 369 391 C2H2-TYPE.
 ZN_FING 397 419 C2H2-TYPE.
 ZN_FING 425 447 C2H2-TYPE.
 ZN_FING 453 475 C2H2-TYPE.
 ZN_FING 481 503 C2H2-TYPE.
 ZN_FING 509 531 C2H2-TYPE.
 ZN_FING 537 559 C2H2-TYPE.

FT ZN_FING 565 587 C2H2-TYPE.
 FT ZN_FING 593 615 C2H2-TYPE.
 FT ZN_FING 621 643 C2H2-TYPE.
 SQ SEQUENCE 645 AA; 73012 MW; 7459CD140F5AF469 CRC64;

Query Match

Best Local Similarity 22.5%; Pred. No. 76-05; Mismatches 195; Gaps 21;
 Matches 126; Conservativity 59;

QY 180 ILEELRYLQGRQIHOMQTEQICRQVLLIGSIGOTGAPASPSELPGTGAASSTKPLPL 239
 Db 211 VIFPMILLTQHCVYREQAYQCSR-----GQEVPSDSPLL-----HQDTLLGK 255
 QY 240 FSPYKPAQTGKTTASSSSSSSSGAEPPKQAFHLYPHLPSQHPFSVGVGRSHKPTAP 299
 Db 256 KSPYH--STHKDTHSPVPPIQPSVHPGRKRW--CHEGKGF-----RQ 296
 QY 300 SPALPGSTDLIASPHLAPGTTGLIAOCLGARGLEAASPELLPKNSGELGYGV 359
 Db 297 SSAL--QTHQHV-----HTGEKPYRCDSCGKGF-----SRSSDLNIHR 333
 QY 360 ISSLEKPGRKRCARCAVFGSDSALQIHLSHTGEPYKCNVCNRPRTTGNLKVHFR 419
 Db 334 VHTGEKP---YKCEYCGKGFQMAHLQAHRIHTGEKPYKCGDKRSCSNLTHQV 390
 QY 420 HREKYPVQMPNHPVPEHLDVYITSSGLPYGMS-----VPPEKAEERAGTPGG 467
 Db 391 HTEKRP-----YECNEGKRPSTLSGNDIHQVHTGEKPYKCE-----CGK 432
 QY 468 GVERKPLVASTTALSATSLTLSTGTSTAVAGLPYFNKFLMKAVPKSADENTPPG 527
 Db 433 GF-----SSASSFQSHQVHTGEK-----PFH----- 454
 QY 528 SEGSAIAGVADSGSATRMQSLKVTLS-----LPSMALITNHLKSTGSEPPPYV 575
 Db 455 ---CSVCKNFSRSSHFLDHQRIHTGEKPYRCVCGKRP--WLSLSHSHQSVHTGKKPYK 510
 QY 576 LEPLGASPSSTSKLQOLYKRIDRQGAVAVAASGAPTTAPADSSASGPNOCYICLV 635
 Db 511 CGEGCKGFSHMSLQ-----AHHSVHTG-----EKPRCNVCORQ 545
 QY 636 VLSCPALRLHYGCHGGRPFCKYCGRAFSTRGNLRAH---FVGHKTSPPARAONSCPIQ 692
 Db 546 FSKTSNLOAHORVHTGEKPYKCDTCGKAFSOKSSLQYHQRITHTEKPP-----KCEEC 598
 QY 693 OKKFTNAVTLQOHVHMHG 711
 Db 599 GKEFRWSYGLSHQVHTG 617

Search completed: January 13, 2003, 15:19:06
 Job time : 37.4828 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 13, 2003, 15:17:45 : Search time 42.9357 Seconds
(without alignments) 4808.565 Million cell updates/sec

Title: US-09-988-117-3
Perfect score: 5301
Sequence: 1 MA0ETGSSSRLLGPGCEPAE.....SSIPSPGLSPPEPRKDDPTMP 1002

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database :
1: SPREMBL_21:*
2: sp_archaea:*
3: sp_bacteria:*
4: sp_fungi:*
5: sp_human:*
6: sp_invertebrate:*
7: sp_mammal:*
8: sp_mhc:*
9: sp_organelle:*
10: sp_phage:*
11: sp_plant:*
12: sp_rodent:*
13: sp_virus:*
14: sp_vertebrate:*
15: sp_unclassified:*
16: sp_rv1rus:*
17: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1265	23.9	1350	13	Q91929
2	1260.5	23.8	1377	13	Q9DDN5
3	1258.5	23.7	1261	13	Q802M5
4	1255.5	23.7	1308	13	Q9DF77
5	1165	22.0	1272	11	Q9EPW7
6	1142.5	21.6	1061	13	Q9PVN7
7	1055.5	19.9	826	13	Q9PVN6
8	1014.5	19.1	898	13	Q12958
9	1007.5	19.0	1053	4	Q9U0Q4
10	915.5	17.3	382	13	Q9W7G2
11	765	14.4	1373	5	Q9VKH2
12	742.5	14.0	1263	5	P91639
13	736.5	13.9	1267	5	Q9VKH3
14	663	12.5	299	13	Q9DGH3
15	654	12.3	327	13	Q9DGH1
16	627.5	11.8	288	13	Q9DGH2

17	526	9.9	549	13	Q9PVN5	Q9PVN5 xenopus lae
18	377.5	7.1	744	5	Q17396	Q17396 caenorhabd1
19	349	6.6	2282	11	Q61479	Q61479 mus musculu
20	345	6.5	2406	4	Q9B2S0	Q9B2S0 homo sapien
21	337	6.4	927	4	Q9UEG4	Q9UEG4 homo sapien
22	336.5	6.3	1300	4	Q15090	Q15090 homo sapien
23	330.5	6.2	619	4	Q96K58	Q96K58 homo sapien
24	330	6.2	1173	13	Q90783	Q90783 homo sapien
25	325.5	6.1	1258	4	Q9P2A7	Q9P2A7 homo sapien
26	322.5	6.1	744	5	Q24219	Q24219 drosophila
27	321.5	6.1	498	4	Q9H8L4	Q9H8L4 homo sapien
28	320.5	6.0	751	4	Q9NS43	Q9NS43 homo sapien
29	315	6.0	1615	13	Q57415	Q57415 gallus gall
30	313.5	5.9	1237	11	Q9D2D7	Q9D2D7 mus musculu
31	312	5.9	734	4	Q96171	Q96171 homo sapien
32	311.5	5.9	744	5	Q9V0U9	Q9V0U9 drosophila
33	311.5	5.9	1920	5	Q46205	Q46205 drosophila
34	310	5.8	962	5	Q61360	Q61360 drosophila
35	309	5.8	736	5	Q960L6	Q960L6 drosophila
36	309	5.8	934	5	Q9YKF1	Q9YKF1 drosophila
37	309	5.8	1893	5	Q9W4J1	Q9W4J1 drosophila
38	307.5	5.8	1891	5	Q77275	Q77275 drosophila
39	304	5.7	725	4	Q96S24	Q96S24 homo sapien
40	300.5	5.7	644	4	Q9NT61	Q9NT61 homo sapien
41	300.5	5.7	823	4	Q9H162	Q9H162 homo sapien
42	300.5	5.7	2232	5	P91365	P91365 caenorhabd1
43	299	5.6	561	4	Q96BV0	Q96BV0 homo sapien
44	299	5.6	894	4	Q9C0K0	Q9C0K0 homo sapien
45	296.5	5.6	654	4	Q96CP9	Q96CP9 homo sapien
46	296.5	5.6	754	11	Q99J65	Q99J65 mus musculu
47	296	5.6	1104	4	Q9P243	Q9P243 homo sapien
48	295.5	5.6	563	11	Q921H7	Q921H7 mus musculu
49	295.5	5.6	591	11	Q8R0V0	Q8R0V0 mus musculu
50	294.5	5.6	700	11	Q99K53	Q99K53 mus musculu
51	294.5	5.6	701	11	Q9R164	Q9R164 mus musculu
52	294	5.5	835	4	Q9H165	Q9H165 homo sapien
53	292.5	5.5	812	11	Q99PV7	Q99PV7 mus musculu
54	292.5	5.5	995	11	Q35615	Q35615 mus musculu
55	292	5.5	624	11	Q62065	Q62065 mus musculu
56	292	5.5	726	11	Q924S6	Q924S6 mus musculu
57	291	5.5	783	11	Q9QY56	Q9QY56 mus musculu
58	290	5.5	814	11	Q70162	Q70162 mus musculu
59	289.5	5.5	784	4	Q00146	Q00146 homo sapien
60	289.5	5.4	884	11	Q99PV8	Q99PV8 mus musculu
61	288.5	5.4	662	5	Q9N360	Q9N360 caenorhabd1
62	287.5	5.4	596	11	Q9CWF3	Q9CWF3 mus musculu
63	287	5.4	813	11	Q9JLF8	Q9JLF8 mus musculu
64	286.5	5.4	1173	11	Q63624	Q63624 rattus norv
65	286.5	5.4	1190	4	Q99621	Q99621 homo sapien
66	285.5	5.4	567	4	Q99W60	Q99W60 homo sapien
67	285.5	5.4	567	4	Q9BU26	Q9BU26 homo sapien
68	285.5	5.4	567	4	Q9BU26	Q9BU26 homo sapien
69	285	5.4	1606	11	Q924A2	Q924A2 mus musculu
70	283	5.3	1453	4	Q9Y6T1	Q9Y6T1 homo sapien
71	282.5	5.3	567	4	Q9NZH2	Q9NZH2 homo sapien
72	282.5	5.3	841	11	Q9WV18	Q9WV18 mus musculu
73	281.5	5.3	756	5	Q61362	Q61362 drosophila
74	281	5.3	1608	4	Q96RK0	Q96RK0 homo sapien
75	280	5.3	703	11	Q9R161	Q9R161 mus musculu
76	279	5.3	650	11	Q91WF9	Q91WF9 mus musculu
77	278	5.2	744	11	Q62788	Q62788 rattus norv
78	278	5.2	3942	11	Q88737	Q88737 mus musculu
79	277.5	5.2	568	11	Q8R0R2	Q8R0R2 mus musculu
80	277.5	5.2	599	4	Q9F2F9	Q9F2F9 homo sapien
81	277.5	5.2	2703	5	Q9VEG7	Q9VEG7 drosophila
82	277	5.2	540	11	Q9JIB9	Q9JIB9 mus musculu
83	277	5.2	618	4	Q96MX3	Q96MX3 homo sapien
84	277	5.2	2715	5	Q61603	Q61603 drosophila
85	275.5	5.2	977	4	Q96H84	Q96H84 homo sapien
86	275	5.2	581	11	Q9ERR8	Q9ERR8 mus musculu
87	275	5.2	743	11	Q9R5D1	Q9R5D1 mus musculu
88	275	5.2	1312	4	Q9NR59	Q9NR59 homo sapien
89	273.5	5.2	556	4	Q9UMP5	Q9UMP5 mus musculu

[illegible]

Oy	537	ADSGATMQLSKLVTSLPSMALTLNHLKSTGSPFPVYLEPLGASPSFEKSLQOLVEKI	596
Db	702	TDSGIS--VMFNNLL--LP---LMSQIFA--KFFPGGLDVTIPA--SEESKLQOLVENI	750
Oy	557	DROGAVAAVASTAGAPPTTSAPAPSSSASPNOCVICRLVLSCPRALRLHGOHGGERPFK	656
Db	751	DKK-----SSDPNECIVCHRYLSCOSALKMHYRTHTGGERPK	787
Oy	657	KVCGRAPSTRGNLRANHFVGHKTSPAARAKNSPCTOKKFTNAVVTLOQHRLMLGGIIPN	716
Db	788	KVCGRAPFTTKGNLKTHYSVHRAMPPLRVQHSCTPCOKKFNTNAVVLQGIIRMGHGGIIPN	847
Oy	717	GGSALSEGGAOENSSSEQSTAAGPSGFPOPOSQOPSPEEEMSEEEDEEDEEDVD--	774
Db	848	TPYA-----ENRPD-SMGSDIGSF-----DETIDDLNFSDENNEDCDPDS	888
Oy	775	-----BDSLNG-----RGSSEGEKAI	791
Db	889	VDPTRKSIDASODSLSSSPLEVVSIITALENMOKLINAGLAEQLOASIKSAENGSVEGD	948
Oy	792	SVRGSPSEVS-GAEENVAVSVAAPTIVKEKDENSEKRAPQHTLPPPPPDDMDLDHPQREOG	850
Db	949	GMTUNSSSSIGCMESQSASPAASESTYSN-----TALSPPNSTIDYLKSPNTEK-	999
Oy	851	TSADVSGAMEEEKLEGIS--SPMAALTQEGESTFVLEELNI.PEAMKKPDGESSG----	904
Db	1000	-----LORAVSLDPTNLGSLPIPA-----NGALDLITS-NIDKYIKEEP---LYVLFP	1043
Oy	905	-----RKACEVGCGSFPPTQTALAEHQTKPRKDGPLEFTVCFRCROGLDRATLKKHML	956
Db	1044	FDRCKRYKKNCTIDICKGTACQCALDIHYRSHTKERP-FICTYCNRGFSFKGNLKHMLT	1102
Oy	957	AHHQV-----PPFAHQPN-IATL-----SLVRS-----	980
Db	1103	--HOMDLPQLSEFPSSSMTPNPPTIPSAPSNPATIITKTFENGFMHSSODIKEQPPTNI	1160
Oy	981	-CSSSIPSPGLSP	992
Db	1161	VSSGSLPSSATS	1173
 RESULT 3 ID 090ZWS PRELIMINARY; PRT: 1261 AA.			
AC	090ZWS:		
DT	01-DEC-2001 (TREMBLrel, 19, Created)		
DT	01-DEC-2001 (TREMBLrel, 19, last sequence update)		
DT	01-JUN-2002 (TREMBLrel, 21, last annotation update)		
DE	Spatl 2 protein.		
GN	SAIZ.		
OS	Gallus gallus (Chicken).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
CC	Archosauaria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;		
OC	Gallus.		
OX	NCBI_TaxID=9031;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RA	MEDLINE=21184122; PubMed=11287198;		
RX	Farrell E.R., Tosh G., Church E., Munsterberg A.E.;		
RT	"Cloning and expression of CSAIZ, a new member of the spatl gene		
RT	family in chick."		
RL	Mech. Dev. 102:227-230(2001).		
DR	EMBL: AF304358; AAK38370.1; --		
DR	InterPro: IPR000345; Cytochrome_bind.		
DR	pfam: PF000096; zf-C2H2; 7.		
DR	PROSITE: PS00190; CYTOCHROME_C; UNKNOWN_1.		
DR	PROSITE: PS00028; ZINC_FINGER_C2H2_1; UNKNOWN_7.		
DR	PROSITE: PS50157; ZINC_FINGER_C2H2_2; 7.		
KW	DNA-binding; Zinc-finger.		
SEQ	SEQUENCE 1261 AA; 134317 MW; BB22E257BCAEB565 CRC64;		

Query Match 23.7%; Score 1258.5; DB 13; Length 1261;
Best Local Similarity 31.0%; Pred. No. 1.9e-66;
Matches 367; Conservative 141; Mismatches 341; Indels 333; Gaps 39;

16 GEPARGGDASE-----EHHQVCAKCAQFSDPTFEFLAHQNSCCTDPPVAVI-----I 64
16 GEGADGGGSGNSRSGSEETNVCCECAEFKWTDFLEHKRCKTKNPLVILIVDEDAAP 89
65 GGOENPSSNSASAPREGHRSOVMDEHSNPPDSSGSPDPMTGPERGSESSQGL 124
90 PAEEPPPEPPASS---PSDAQESA--AEEGVQPENSSESSEKSTEKEEPEMEVETAEKS 144
125 VAATGTAAGGGGGLIASPKLATPLPESTPAPPPPPP----- 165
145 FQNGSTN-----TATPLPOLPEPSPWTXYTMTNTVLTLELLSTKVAAQF 191
166 -----PPPGVSG--HNIPLILEELRVLQORIHOMQMTEDQICROYILLG-----S 210
192 SOSARSAASASISGVTAVAIPIMLEQLMALQOQOIHQQLLEQIRSVYAMNRQLRPA 251
211 LGQTVGAPAPSELPPTGAAS-----STKPLPLPSPIKPAOTGKTT----- 252
252 LNPVVAAGV-----TGQASNLQGFATSAIQLTAVLPRAIMQAAGAPPAFDGSH 305
253 -----ASSSSSSSSGAEPKQAFPHLYHPLGSHPSVG----- 287
306 ISRPTSGASTPNISSGSSAPRESSAPCSSNAITSVTPVSVNTTISAQPNASTPSSI 365
288 GVGRSHKPTAPSPPLPGSDOLIASPHLAPRGTGGLAACLGAARGLAASPGLLK 347
366 GHGSLTISLNDPLPQFOT-----SSNSVIFPNPLVSTIA---TANALDPLSA--LMKH 414
348 KNGS--GELGYEVLISLEKPGRHKRCFCACVFGSDSALQIHLRSHTGPERKYCNVGNR 406
415 RKGRPNVSVFEPKSSSEDPFEKHKRCFCACVFGSDSALQIHLRSHTGPERKCNICGNR 474
407 FTTGNLKVHFRHREKTYHVQNPVPHVPHLYITSSGLPYGASVPEK-----AEE 461
475 FSTGNLKVHFRHREKTYHVQNPVPHVPHLYITSSGLPYGASVPEK-----AEE 461
462 AGTP-----GGGVKRPVLAIVSTALSTATESITLS-----TGSTAVAPGLPTENKF 508
535 PVLPTIPISIGLQLEPTTIGVNSYGDSPSTTPMRSRSPORPSASSECTSLSLNTESEG 594
509 VLKMAVERK-----SKADE--NTPPGSE-----GSAIAGVADSGSATRMLSKLYT 552
595 VPVASESPQVQSSSVTKAEPISLPASTRLGDHSLSGYSTASTSSPIITYTDSVST 654
553 SLSPWALLTNHLKTSGFPPVYLEPLGASPETSKLQOLYEKIDRQCAVAVASTAGAP 612
655 SLPPVPLPAVSDQFAKFFFGGLDSM--QTSSTKLOQLEVENIDRK----- 699
613 TTSAPAVSSASGPNQCVLCRLVSCPRALRLHYGOGHGEPEFKCVGRATRNGLRA 672
700 -----MTDPNOCVICHVLSQOSALKMHYRTHGEPEFKKICGRATTKGNLKT 749
673 HFVGHKTSPPAARAQNSCPIQCKFTNAVTLOQHVRLHGLGQIPNGSALSEGGA--- 728
750 HFVGHKTSPPAARAQNSCPIQCKFTNAVTLOQHVRLHGLGQIPNGSALSEGGA--- 728
729 -----QENSESGS-----TASGSGSPQPOQOPS 753
808 LSTDEKNDITLSNFDIDIDENSEDEPELKDMSADSAKPLIYSGSCSPSPSVYISSIA 867
754 PEEEM-----SEEEDEEEDVDEDESLAGRG---SESGEKALISVRGDS 797
868 LBNQMKMIDSVNCOQLSLKSTENGSGESDHLSDSSSANGDLESQASNGSPAMS----- 922
798 EEVSGAEEVATVAAPTIVKEMD-----SNEKAPQH---TLPPPPPPDNLDH 843
923 -----ESSSSMQALSPVNSSEFSRKSQGLSNQDEPQEIOLKTEKPDSPPP----- 969
844 POPMEQG-----TSVSG--AAEEEAKLEGISSPMAL---TQEGEGISTPLVEELNLPE 893

Db 970 --ATENGCALDLTSTNPGRPVIEEA-----PFSLLFLNREKRPSSQSTPLVSTAPT 1020
QY 894 AMKD-----PGSSGRKACEVCG 912
Db 1021 MIKMGVNSHKSPISLGEVPSLPAIGVAPAPQIVMSPGITPMLAPPRTPRQHNQSCG 1080
QY 913 QSPFQATALEEHQKTHPKDQPLTFVFCROGFLDRATLKHM 954
Db 1081 KTFSSASALQIHERHTHGKRP--FGCTTCGRATTKGNLKVHM 1121

RESULT 4
09DEF77 PRELIMINARY; PRT: 1308 AA.
AC 09DEF77;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Spalt 1.
GN SAL1.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RX SEQUENCE FROM N.A.
RA MEDLINE=20442162; PubMed=10985862;
RT "Faircliff E.R., Munsterberg A.E.;
RT "Chick spalt 1 gene expression is controlled by a combination of FGF
RT and Wnt signals in developing limb buds."
RL Dev. Biol. 225:447-458(2000).
DR EMBL: AF288697; AAC13011.1;
DR HSSP: p15822; IBSO.
DR InterPro: IPR000345; CytoC_heme_bind.
DR InterPro: IPR000822; ZnF_C2H2.
DR Pfam: PF00096; Zf_C2H2_10.
DR ProDom: PD000003; ZnF_C2H2_1.
DR SMART: SM00355; ZnF_C2H2_1.
DR PROSITE: PS00190; CYTOCHROME_C; UNKNOWN_1.
DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; 9.
DR PROSITE: PS50157; ZINC_FINGER_C2H2_2; 9.
KW DNA-binding; Zinc-finger.
SQ SEQUENCE 1308 AA; 139846 MW; 81810CC16336298E CRC64;

Query Match 23.7%; Score 1255.5; DB 13; Length 1308;
Best Local Similarity 30.1%; Pred. No. 3e-66;
Matches 359; Conservative 148; Mismatches 320; Indels 365; Gaps 42;

QY 19 AERGD-----ASEHHNPQVCAKCAQFSDPTFEFLAHQNSCCTDPPVAVIIGGOENP 70
Db 22 SQRNGDLEKGGGNRTTKKDAHVCGRCAEFELSDLOHKRCKTKNQLVLI---NENP 78
QY 71 SNSASSAPREGHRS-----RSQVMDTEHSNPPDSSGSS 103
Db 79 ASPPEPTPPRSPSPNDPEQMDYVNTDQVDCSDLSEKRLKDESMDEFASSINNSSSS 138
QY 104 GPDPPTGPERGSESSQGLFVAATGTAG-----GGGLI----- 139
Db 139 -----SKSVNNSTISSNSSTMGTSVTTSLPHIGDITLTIGNFSVINSNVITE 185
QY 140 -LASPKLATPLPESTPAPPPPPPPPPPPVGSGHILNPLILEELRVLQORIHOMQMT 198
Db 186 NLQSTKVAVAQFSGEAR-----CNGASNNKLAVPALMQGLALQOQOIHQQLI 234
QY 199 EQICROYVLGSLQCTVGAPAPSELPPTGAASSTKPLPLPSPIK---PAQTKTQASS 255
Db 235 EQIRHOILLIAS--QNTDMPITSSSPQGLTRASA--NPLSTLSHLSQOLAAAGLAQSLA 291
QY 256 SSSSSSSGAE--PPKQ-----AFHLYHPLGSHPSVGVG 290
Db 292 SOSASISGVKQLPPIQLPQSNPGSTLIPSSSGSSPNINILAAVTTTPSSSEKVASSTGSSQ 351

Y	630	VICLVVSCSPRALRLHYGHGGERPFCVKVCGRAESTRGNLRAHFVGHKTSPPARAONSC	689
Y	658	ICHVILSCOSALKMHYRTHTGERPFCCKICSRDFTTKNLKTHSVHRAAMPRLVQIISC	717
Y	690	PICOKKPTAAVTLQOIHVMILGGIIPN-----GGSLISEGGGAQEN-----SSEQ	735
b	718	PICOKKFTFPA-CIPAAIYRLHMGSOIPNTPVPDNYPEISMESDTSQFDEKNPDDLDNFSDEN	776
Y	736	STASPGSPPO--POSOOPSPPE-----EMSE-----EEE	763
b	777	MECEPESGIPDPFKSADASQDSISSPPLFLEMSTIALENQKMINAGLADQQLQASLSKY	836
Y	764	EDEEEEDVTDDESLAGRGSGSGEKALSYAGDSEEVSCAEEEVATVSAAPTYKENDSN	823
b	837	ENGSMGVDVLNDS-----SSVGGDMISQSGSPAISSSTISQALSPSNST	883
Y	824	E---KAPQHTLPPRPPLPDLNH---POPMQGTSDVSCAMEBEAKLEGISPPAALTOE	877
b	884	QEFHKSFGMEKQRQVRGPGEFANLSTPTPVNAGLALDITSNAEK-----LIKED	932
Y	878	GEGTSTPLVEELNLPRAMKDDPGESSGKACEVCGSQSPFTQALAEHOKTHPYKDGPLETC	937
b	933	SLGILFPF-----RDRGFKF-NTACDIDICKETFCQSGALDIHYRSHTKERP-FIC	979
Y	938	VFCRQGFDRATLTKHMLLHNHOV-----PPRAPHQNIATLSLVGCGSSITSP-	988
b	980	TVCNNGFSTRKGNLKHQMLLT--HQMRDLPSQLFEPBSNIGFGNONSAYIIPANSLSLITFV	1037
Y	989	----GLSPPRKDDPT	1000
	1038	NGFVHVSPPQSDSKAPT	1053

9PVT7
D Q9PVT7 PRELIMINARY; PRT; 1061 AA.

09PVTN; 01-MAY-2000 (TREMblrel. 13, Created)
TT 01-MAY-2000 (TREMblrel. 13, Last sequence update)
TT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
E Xsal-3 long form.
S XSAL-3.
S Xenopus laevis (African clawed frog).
C Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
C Amphibia; Batrachia; Anura; Mesobatrachia; Pipiloidea; Pipidae;
X xenopodinae; Xenopus.
NCBI_TaxId=8355;
[1]

SEQUENCE FROM N.A.
MEDLINE-99456836; PubMed-10527856;
Onuma Y., Nishitakamura R., Takahashi S., Yokota T., Asashima M.;
"Molecular cloning of a novel Xenopus spalt gene (Xsal-3)." ;
Biochem. Biophys. Res. Commun. 264:151-156(1999).
EMBL; AB030827; BAA85900.1; -.
HSSP: P15822; IBOO.
InterPro: IPR000822; Znf_C2H2.
Pfam: PF00096; zf_C2H2; 8.
ProDom: PD000003; Znf_C2H2; 1.
SMART: SM00355; Znf_C2H2; 7.
PROSITE: PS00028; ZINC_FINGER_C2H2_1; 7.
PROSITE: PS50157; ZINC_FINGER_C2H2_2; 7.
DNA-binding: Metal-binding; Zinc-finger
SEQUENCE 1061 AA; 114303 MW; 7C5D6289CCFA864B CRC64;

Query Match	21.6%;	Score 1142.5;	DB 13;	Length 1061;
Best Local Similarity	30.6%;	Pred. No. 1.2e-59;		
Matches 323;	Conservative 133;	Mismatches 52;	Indels 249;	Gaps 33

```

2 AQETGSSSRLLGGPCGEPAERGDASEEHHPQCAKCAQAFSDPTEFLAHQNSCCTDPPVM 61
  :: | : : - | | : | | | | : : : | : | : | : | : | : | : | : | : | : | :
14 SEEDASAOCKDCPSSYDGERAVKRCRMEETHICEKCCABFFELSDLEHKKSCSTKTPVL 73

```

QY	62	VILGGQENPNSSASSAPRPGCHSNOYMDTHSNRPDSSGSGPPRTMGPERGEESSG	121
Db	74	IMDNG-BGAMSHDGYTEESPGEAA-----APSEVPADVQPKRSSVSPERMEKRVDS	126
QY	122	QOFVATAGTAAAGGGGLLAPKLGATPLPRESTPAPRPPPPPPPGVSGSHLN--IP	178
Db	127	NKTYPPGNSKTINGIG--YVFKTNLSNTNVTILOTINSTKVAVNOHASDGAITSNTNPAIP	184
QY	179	LILEELRVLOOROIHOMQOIEOICQVILL--GSLGQTVGAPASPELPGTGAASSTKPL	236
Db	185	IILEQVLCLOQOQOIQOIQLTQEQIRQIOMAMAPNSLHPSTAAATDPLKALGAHLSQOLSA	244
QY	237	LPLFSLPKPAQTKTITASSSSSSSSSSGAEPKQAF--PHLYPLGSGHPFVSGVGGRSEK	294
Db	245	VALIGQAKGQTSLESLESKOBLPHSNVAMPACTVPLPLTTLTKOEPR---NLGLITNA	300
QY	295	PTPAPSPALPGSTQQLIASPLHAPFGTGTGLAAGTGAARGLEAASPGILKPKNGSEL	354
Db	301	VGRFPNPAIHS-----PGT-----ITIPNPINALDSPSKKLKYK-----	344
QY	355	GXGEVYISLEKPGG---RHKRCFCAKVFQSDALQIHLRSHTEGRPYKCNVGNRFTT	409
Db	335	-PFTVTPPEAKPGNEDQLEFRHKCFKCFKGVFGDSALQIHLRSHTEGRPYKCNICGARFTT	393
QY	410	RGLUKYAFHHHREKYPHYQOMPHVPEHLDYITSSGLPYGKSVPEKKEDEEAGTGGGV	469
Db	394	KGNLKVHFOHKKDKYPHIKMYPVPEHLDVNPPTSGIPLYGSMVPLDESLADTSG--	451
QY	470	ERRPLVASTALSTESITLLSTGSTVAVQGLPFFNFKVLKAVEPKSKADENTPGESE	529
Db	452	-----LGLPBPATNIS-GLITEVLAQAPL-----NMOSRBPGESE	485
QY	530	GSAI--AGVADSGSARMOJSLKVTSLPSWALLTNHLKSTGSPFPYVLEPLGASP---	583
Db	486	GESVSSGAVQESG-----TDQSFNSP---PVSGSSEGG	516
QY	584	SEISKLOQLEYEKIDROCAVAVASTASGAPTTSAPRPSASGSPNOCYTLRYLSCPRALR	643
Db	517	SETTKLOQLEVENEDKNK-----SETNCLLICHRYLSCPPSLIK	553
QY	644	LHFGQHGGEPPKCKVCGRAFTSTGRNLRAHVGHKTSPAARONSCPTIOKFFTNVTLQ	703
Db	554	MHYRTHTGERPFCKICIGRAFSTSKNLKTHGVIRANPLKLOHSCPLIOCKKFTNAVILQ	613
QY	704	QHVHMLGGQIPNG-----GSALSEGGAQENSSEOS-----TASG	740
Db	614	QHIRMHMGKILPMTPVSEASDDIDSMDEKNGELNNSFTDENLDDIMEDELEMANAG	673
QY	741	PGSFPQPOSOQPSPEEEMSEEE---EEDDEEDVDDEDSL-----AGRSSESQGEK	789
Db	674	-----SKPFPHSTRAESPAMQFSTGGODKPVLPALSNALQRONSVKSENGSLE	725
QY	790	AISVRDSEEVSGAEEEVANVSVAFT-----TYKEMDSNEKAPQHTLPPPPPPPN	840
Db	726	SQGLINDSSVMDQETFGKS---PTQSEARTFSPNTSOSDSMAS-----KSPSYNG	775
QY	841	LD-----HPO-----PMEQSTSVS-----GAM	858
Db	776	LDDLGMLSKREHSONGSLNDDGQALDLITNGCFARKIKEPFGIHLONGEFGRLPNLYGAP	835
QY	859	EEBKALEGISSPMAALTOE--GEGTSTPLVEELNLPEAMKDKDPRESSGGRACVCGQSFTT	917
Db	836	PAIKMEVSSDRMAGATQYLGPRMLSPGLNPLIYFOR---RSAKQHCITMCKKNFSS	889
QY	918	QTALEHOKTHPKDGLFTVCVCGQGGILDRATLKKHM	954
Db	890	ASALQIHERHTHEKPR-FACTICGRAFTTKGNLKVHV	925
RESULT	7		
AC	Q9PVN6	PRELIMINARY;	PRT; 826 AA.
AD	Q9PVN6;		

Q9PVN6	PRELIMINARY;	PRT;	826 AA.
ID Q9PVN6			

Q9PVN6; AC

DT	01-MAY-2000 (TREMBLrel. 13, Created)
DT	01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT	01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE	xsal-3 short form.
GN	XSAL-3.
OC	Xenopus laevis (African clawed frog).
OC	Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi:
OC	Amphibia: Batrachia: Anura: Mesobatrachia: Pipidae: Pipidae;
OC	Xenopodidae; Xenopus.
OX	NCBI_TaxId=8355;
OX	NCBI_TaxId=8355;
RP	SEQUENCE FROM N.A.
RP	MEDLINE=99458636; PubMed=10527856;
RX	Onuma Y., Nishinakamura R., Takahashi S., Yokota T., Asashima M.,
RA	"Molecular cloning of a novel Xenopus spalt gene (xsal-3).";
RT	Biochem. Biophys. Res. Commun. 264:151-156(1999).
RL	EMBL: AB030827; BAA85901.1; -
DR	Interpro: IPR000822; Znf_C2H2.
DR	Pfam: PF00096; Zf-C2H2; 6.
DR	Prodom: PD000003; Znf_C2H2; 1.
DR	SMART: SM00355; Znf_C2H2; 5.
DR	PROSITE: PS00028; ZINC_FINGER_C2H2_1; 5.
DR	PROSITE: PS0157; ZINC_FINGER_C2H2_2; 5.
KW	DNA-binding: Metal-binding; Zinc-finger.
KW	SEQUENCE 826 Aa: 89219 MW, BF76B5F58C9B63D CRC64;
Query Match	19.9%; Score 1055.5; DB 13; Length 826;
Best Local Similarity	30.9%; Pred. No. 1.2e-54;
Matches 296;	Conservative 122; Mismatches 328; Indels 211; Gaps 30;
QY	2 AOEFGSSRLGRCPCGERAERGSDASEENHPOYCAKCAOFSDFTEFLAHONSCCTDPYV 61
DB	14 SEEDASARCKDPPSSVDGERAVKRCMEETHICEKCAEFELSDLEHKKSCKTPVL 73
QY	62 VIIGGEONSNSASSAPREGRSROVMDEHNSHPROGSSGPRDPTMGPERKESG 121
DB	74 IANDG-EGMASHDGTIEESPGEAA-----APSEVPAVDSDQKSSVSPEKKEKYDS 126
QY	122 QFLVATGTAAGGGGGLIASPKLATLPRESTRAPRRPPRRPGVSGHLN--IP 178
DB	127 NKTVPQNSKTNGLG--YVFKTNLSNTNVTLOTINSTKVAVNQHASDGVATSATNPAP 184
QY	179 LILELRLVLOQRIHOMOMTEQICROYLL--GSLGQVYGARASBELPRTGAASSTRPL 236
DB	185 IILEDLVLCQQOQLQOIOLTEQIRIQIAMAPNSLHPSTIAATDPLKALAHLSQOLSA 244
QY	237 LPLFPIKRAQIGKTTASSSSSSSSSGAEPKQAF--FILNPLGSHPRPSVGSGRSH 294
DB	245 VALTIQKAGTOSLSELSKOSKLRPHSNVAMPAGTVPALTLTSLKOE--NLGLTNA 300
QY	295 PTPAPSPALPGSTDQILASPHLAFPGTGLLAQCLGARGLEAAMASPGILKPKNGSEL 354
DB	301 VGRFPNAPLPHS-----PGT-----IIPGNPINALDPSKTLKYV----- 334
QY	355 GYGEVYISLEKPGC-----RHKCRFCARFEGSDALQIHLRSHTEGERPYCNVCGNRETT 409
DB	335 -FRPTVTPPEAKPQENEDLFRHKCKFCGCVFGDSALQIHLRSHTEGERPYCNVCGNRETT 393
QY	410 RGNLKVHFRHREKYPVONMHPVREHLDVITSSGLRYGVMPVRPKAEAEACTPGGCV 469
DB	394 KGNLKVHFRHREKYPVONMHPVREHLDVITSSGLRYGVMPVRPKAEAEACTPGGCV 451
QY	470 ERKPLVASTTASATESLTLSTGTSTVAVAGCLPTFNKFLMKAVEKSKADEWTPPGE 529
DB	452 -----LTGLRPSATNLS-GLTEVLAQFPL-----NMQNSRPSGE 485
QY	530 GSAI---AGVADSGSATRMQLSKLVSLPSMALTLNLHLSSTGSGFPPEYVLEPLGASP--- 583
DB	486 GESVSSGAVQVEEG-----TDQSFNSP-----PVSGSSSEG 516
QY	584 SETSKLOOLVEKIDROGAVNAVASTAGAPPTTSAPAPSSSASGPNOCYICILVGLSPRLR 643
DB	517 SETSKLOOLVEKIDKNS-----SETNECLICHTVILVGLSPSSLK 553

Oy	644	LHYGGGGERPFCKCYKCGRAFTSGNLRARHVEGHKTSPPARAQNOSCPIQCKRTNAVTLQ	703
Db	554	MHYRTGTGERPFCKCICKGRAFKSTSKNLTKYHVHRANPPLKIQHSCPICQKRTNAVYLQ	613
Oy	704	OHVRHLGGGIIPNG-----GSALSEGGAQENSEOS-----TASG	740
Db	614	QHIRMAMGKITPTPVSEASDDIDSMDERKNGLNSTDENLDDIDMEDDELAEANASG	673
Oy	741	PGSFPOPOSQPSPBEEMSEEE-----EDEDDEEDVDDEDL-----AGRSESQGEK	789
Db	674	-----SKRPPIPHSTRAESPAMQFSTGQGDKVPVLPSALNLNQRONSVKSSENGSL	725
Oy	790	AISVRCDSEEVSGAEEBVATSVAAPTYTYKENDSNEKAPOHTLP PPPPPD-NLDHPME	848
Db	726	SQGLTJLNDSSVWMOQEYPFGKS--PT-----QSEARTSPITSQSNSKSPSPSY	773
Oy	849	OGSDVSQGANEEFEKLEGISSPMALITDOEGESTNPVLELNLEPA-----MKKDPG	900
Db	774	NGLDDIGMLSKDERHSONG-----SLNPDGDA-----LDLTNGCFARKIKEEPG	817
RESULT 8			
ID	012958	PRELIMINARY;	PRF: 898 AA.
AC	012958;		
DT	01-JUL-1997 (TREMBUREL_04, Created)		
DT	01-JUL-1997 (TREMBUREL_04, Last sequence update)		
DT	01-MAR-2002 (TREMBUREL_20, Last annotation update)		
DE	Zinc finger protein sal (Fragment).		
OC	Oryzias latipes (Medaka fish).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;		
OC	Acanthomorpha; Acanthopterygii; Percormorpha; Atherinomorpha;		
OC	Beloniformes; Adrianchthyidae; Oryziinae; Oryzias.		
NCBI_TaxID=8090;			
RN	[1]		
RA	SEQUENCE FROM N.A.		
RP	Stick R., Koester R., Witbrodt J,		
RT	"Hedgehog signaling activators split at the midbrain-hindbrain boundary		
RT	of fish."		
RL	Submitted (NOV-1996) to the EMBL/GenBank/DDbj databases.		
DR	EMBL: U77376; AAB51127.1; -.		
DR	HSSP: P15822; IBBO.		
DR	InterPro: IPR000822; znf_C2H2.		
DR	Pfam: PF00096; zf-C2H2_7.		
DR	SMART: SM00355; znf_C2H2_7.		
DR	PROSITE: PS00028; ZINC_FINGER_C2H2_1; 7.		
DR	PROSITE: PS0157; ZINC_FINGER_C2H2_2; 7.		
KW	DNA-binding; Metal-binding; Zinc-finger.		
FT	NON_TER		
SO	SEQUENCE		
	898 AA; 96668 MW; 7BB1CF5D7A104287 CRC64;		
Query Match			
	Best Local Similarity	19.1% Score 1014.5; DB 13; Length 898;	
	Matches 266; Conservative 110; Mismatches 192; Indels 237; Gaps		
Oy	327	AACCGARGGLEAAASPGLLKPKNGS-GELGYEVISLSLEKPGRHKCRFCAKVGSDSAL	385
Db	4	ASTAATNALDPLSA-LMKHKRGKRPANVSVDTPKRSSDEPRFKHKRCFAKVGSDSAL	61
Oy	386	QHILRSHTGERPYKCWCNGNRFTTRGNLKVHRHREKKYRHVMNPMPVPEHLDIVYTSS	445
Db	62	QHILRSHTGERPYKCNICNGNRFSTKGNLKVHFQRHEKYPHIOMNPYPVPEYLDNVPTSS	121
Oy	446	GLPYGMVPEPK-----NEEEAGTGCGGVGR	471
Db	122	GIPYGMSLPPEKPVTTWLDSKRVLPVTPVSVALQLDPTLPSTMIGGFESPSLTPLSRSPQ	181
Oy	472	KPLVASPTTALSATESILTISTGSTAVAP-----GLPTFNKFVLKAVEPSKADENT	524
Db	182	RHSPPSSDECASLSRNVAADSTKTITTSFSNNPNLGSQGPPLLKFBGILLSPVYSARPENT	241


```
RESULT 10
ID 09W7G2 PRELIMINARY; PRT: 382 AA.
AC 09W7G2:
DT 01-NOV-1999 (Tremblrel. 12, Created)
DT 01-NOV-1999 (Tremblrel. 12, Last sequence update)
DE 01-MAR-2002 (Tremblrel. 20, Last annotation update)
DEF Spalt protein (Fragment).
NCBI_Taxid=9031;
Gallus gallus (Chicken).
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauromorpha; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
Gallus.
NCBI_Taxid=9031;
SEQUENCE FROM N.A.
Reveilli J.-P., Thaller C., Etchele G.;
"Evidence for morphogenetic signaling by BMP-2 in the chick limb
bud."
Submitted (NOV-1998) to the EMBL/Genbank/DBJ databases.
EMBL: AF110143; AAD3040.1;
Interpro: IPR000822; Znf_C2H2.
DR Pfam: PF00096; Zf-C2H2; 5.
DR SMART: SM00355; Znf_C2H2; 5.
DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; 5.
DR PROSITE: PS0157; ZINC_FINGER_C2H2_2; 5.
DR KKM DNA-binding; Metal-binding; Zinc-finger.
FT NON_TER 1
FT SEQUENCE 382 AA; 41778 MW; B6622663DEFDE3B6 CRC64;
Query Match 17.3%; Score 915.5; DB 13; Length 382;
Best Local Similarity 47.5%; Pred. No. 1e-46;
Matches 192; Conservative 53; Mismatches 96; Indels 63; Gaps 9;
QY 346 KPNKSGELGIVGENTISLEKPGGRKRCFCAKVGSDALQIHLSHTGERPYKCNVCGN 405
DB 8 KPPN-----VSVEEPKSSSEDPFKKRCRCACAKVFGSDALQIHLSHTGERPFKCNICGN 63
QY 406 RFTTRGNLKVHRRKRYPHYOMNPHRYENDYVITSSGLPYGMSVPRPK-----AEE 460
DB 64 RFTSKGNLKVHRRKRYPHYOMNPHRYENDYVITSSGLPYGMSVPRPKRYTTWIDS 123
QY 461 EAGTP-----GGGVEKKPLVASTATLATESLTLDS-----TGSTVAAPGLPTFNK 507
DB 124 KPLVLTPTSTIGLQRLPTIPGVNSYGDSPSTPMRSRQRPSPASGETSLSPSLNTSES 183
QY 508 FVLKKAVERK-----SKADE-NTPGSE-----GSAIAGVADSGSATRMQLSKLV 551
DB 184 GVPVSAESPPQVQSSSYTKARPIINLPASTRLGDHSLGCGVSTASTSIPTIVDSSVS 243
QY 552 TSLPSMLLTLMHLKSTGSPRPYVLERLGLASPSSTLQQLVLEKIDRGNAAVAATASGA 611
DB 244 TSLPSMLLTLMHLKSTGSPRPYVLERLGLASPSSTLQQLVLEKIDRGNAAVAATASGA 611
QY 612 PTTSAPAPSSASGPNOCVTCRLVSLCPRALRLHYGONGGERPKCKVCGRAFTSGNLR 671
DB 290 -----MTDRNQCICIRVLSQGSALMKHMYRTHTGERPKCKVCGRAFTTKGNLK 338
QY 672 AHFVGHKTSPPARAONSCPTCOCKKFTTNAVTLQOHVNMHLSGOIP 715
DB 339 THFVGHRAKRLRLVQHSCPTCOCKKFTTNAVTLQOHVNMHLSGOIP 382
RESULT 11
ID 09VKH2 PRELIMINARY; PRT: 1373 AA.
AC 09VKH2:
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DE 01-MAR-2002 (Tremblrel. 20, Last annotation update)
DEF Salm protein.
FT SLM OR CG6464.
GN
```

```
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_Taxid=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazer J.G., Champe M., Pfeiffer B.D.,
RA Man K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abriell J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Bailew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cavley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwan C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Mekulov G., Milshina N.V., Moharry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Peled J.M.,
RA Palazolo M., Peltman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Ralner K., Remington K., Saunders R.D.C., Scheefler F., Shen H.,
RA Shire B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spreading A.C., Stapleton M., Strong R., Sun E.,
RA Svirskaas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
DR EMBL: AE003632; AAF53097.1; -.
DR HSSP: P15822; 1BBO.
DR Flybase: FBgn0004579; salm.
DR InterPro: IPR000822; Znf_C2H2.
DR Pfam: PF00096; Zf-C2H2; 7.
DR SMART: SM00355; Znf_C2H2; 7.
DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; 7.
DR PROSITE: PS0157; ZINC_FINGER_C2H2_2; 7.
KW DNA-binding; Metal-binding; Zinc-finger.
SQ SEQUENCE 1373 AA; 151014 MW; A6D857870F645ACC CRC64;
Query Match 14.4%; Score 765; DB 5; Length 1373;
Best Local Similarity 22.9%; Pred. No. 4e-37;
Matches 319; Conservative 151; Mismatches 446; Indels 476; Gaps 46;
QY 3 OETGSSRLRGPGCGEPAERNGDASEHNPVCAKCAQGFDPTEFLAIONSCCTPRVAV 62
DB 50 KDIGSDQENGGCCSPLTATTTASPSRPF-----PEEDPPEEDSTSEQSI---PEQST 100
QY 63 IIGGOENPSNSASAPREGRHSRQVMDTEHSN---PPDSGSSGPPPT----- 109
DB 101 PDHOLENDIKSEAKSLIEPEVDNNNRVAMTKPSSSEERPNMNSGMSPPVAAASAEEAAT 160
QY 110 -WGPERRG-----ESSGQFLVAATGTAAAGGGGLI-----LASPKLGAT 148
DB 161 ERTPEKEKEDEVDVEKPEDEAPSSAVPSTEVTLPGGAGAVTLEAIONMAIAIOFAAK 220
```


Y	149	PLRSTPRRRRRRRRRRRRVGGSHLNIPLLEBLRVLOQROHQMOMTEQICROYLL-	207
Y	221	TIANGSNGADNF-----AAKOLAFLOQPLFNLOQOOLFOIQLLOLOLOLALN	269
Y	208	-----LGSLOQTVGAPASPSLELPTGAASSTPRL	237
Y	270	QAKQEDTEBDADQEDQEDQEDTQYEEERLADMELRQAEARMAE-----AKARONLI	323
Y	238	PLESPILKPAQOTGKTASSSSSSSSSGAERPKQAFNLVYPLRGSQHPFVSGVGSRHKPTR	297
Y	324	NAGVPLR-----ESSGSPASLTKRRRHHDHESQRRPR---SLDTHKADT	366
Y	298	A-----PSRLPGSDP--QLIASPH--LAPRTTGLLAOSLGARGLAASPG	343
Y	367	AODALAKLEMEHTPLRFGSDLASSITTNHDLRPNSLDLOKR---AQEVLDSASOG	422
Y	344	LLK-----PKNSGELGUGEVISLSLEPGGRHKRCFAKVFGSDALQHLNHSHT	393
Y	423	ILANSMAADPFAREKSGEKG-----GRNEPFRKHRCYQSKVUGSDALQHLNHSHT	474
Y	394	GERPKCNVCGNRTTNGNLKVHNRKEXRYHVOYMPNRPVDEH-----LDYVTSS	445
Y	475	GERPKCNVCGSHTKGNLKVHNRQYAKRPHVYMNATPREDHMKPHRLDQMSPTD	534
Y	446	GLPYGMSVPRKEAEEBEGTSG--GVET--KPLVASTALSATESLTLSTGSTAV---	498
Y	535	SSPNHSRPPRLGASAPRPAFRPARGLOLVYPRMEILKISGAAPRQVPELPTDLRKP	594
Y	499	ARGLTENKFLVLMKAVERRKSKADNTPRSGSG-----AIAGVADSSATPRMOLSK-	549
Y	595	SFOLEDREQVKNERVEKQOREHNEOMAESSEDERPRLYLEVRIKEEVEEEOYKOE	654
Y	550	-----LVTSLSWALTLNLKSTGSEFFRVULBEU-----GASPSETSKLOOL---	592
Y	655	DHRIEPRTPRSESHNRSNNHNSHNGYRVUQGIORALNHPGSGSGSHNDHLPTR	714
Y	593	-----VEKDROGAVAVASTASCAPTTSAPR-----	619
Y	715	GOLPREDEFAERPLNFTTAKMLSPREHNSPVRSAGALRPGVRRPNNHNNHMAKSPF	774
Y	620	-----SSASGRPOCYULCVLS	637
Y	775	FNPIKHEMALRPRPSNDNSMNEFIEVSNTCETPKKLELMKNKKTIDPQOCVUCDRVLS	834
Y	638	CPRALRLHNGONGSERPFKCVSGRASTFRCNLRAHFGNKTSPRAAONSCPTQCKKTT	697
Y	835	CKSALOMHNRTHNGERPFKCRIGRAFTTKNLKTNMAVNHKLRPRNRNPFQOCVUCHKYS	894
Y	698	NAVTLQOAHVRLNG-----GQIPNGSALSSEGGAQOENSS	733
Y	895	NALVLOQHRLTHGERTDILPRQIOAELIRPSPMBSGHNPNPFAAFAFHGA----	948
Y	734	EGTASGSPSPRP-----OSQORPREEMSEEBEED-----DEE	770
Y	949	---LPGREGGPRNGAHNGALGSESSQGDMDMDDCGEDVDDVSSSEHLSNSNLQEG	1005
Y	771	D-----VTDESLAGSGSESGEKA---TSVRGDS	798
Y	1006	DRSRGDDOKSLLEBKRLRIDALGCVUVTNIPVRRPSASHSHGVSGSTAPTSPSYANSSQ	1065
Y	799	EV---SGAEEVA-----TVAAPATTVKEMDS-----NEKARQH	829
Y	1066	VIKRSSSPARSEASOGALDLPRAAPRTSSSSSRSLPKKEKVPSPSLPRSPGSSHNASAN	1125
Y	830	TLPPRRP-----PDNLDH-----PDMEQ	849
Y	1126	TLTSELPRVVGIDCLPRGLOHNLQOONHLMQOQAAVAAAAAONHNNHOMALHNOEQ	1185
Y	850	GTSVSGAMEFEKLELGSISSPMAL-----TQEG	887
Y	1186	LRREAAEAOQKAAAAAAAAAAAAAOQOTPRQADQRODQBGSGAPRRPRLMGARRPFGM	1245

```

QY      888  ETLNP-----EAMK-----PGESSGR---ACEYGGSPFOATLE  922
        |||
Db      1246  FPNLPEFPATTQNMCMANMOIQSVAPAPAPFNDLALSGVRSSTCGICTYTFPHCSALD  1305
        |||
QY      923  EHOQHTPKDGLPFTCYFCROGFLDRATLKKMLLAHQVPPFADHPGONITATLSVPGCS  982
        |||
Db      1306  IYRSTTKRRP-FKCSICDRGFTTKGNLKKOHLT--HKINDMEQETFRNNRAVKYV-----  1357
        |||
QY      983  SSIPSHGLSPFP  994
        |||
Db      1358  -SLP---ISPIP  1365
        |||

```

ID	AC	PRELIMINARY	PRT	1263 AA.
P91639				
AC	P91639			
DT	01-MAY-1997	(TREMBLrel. 03, Created)		
DT	01-MAY-1997	(TREMBLrel. 03, Last sequence update)		
DT	01-MAR-2002	(TREMBLrel. 20, Last annotation update)		
DE		Zinc-finger protein SALR (SPALF-related protein).		
GN		SALR OR CG4881.		
OS		Drosophila melanogaster (Fruit fly).		
OC		Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;		
OC		Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;		
OC		Ephydroidea; Drosophilidae; Drosophila.		
OX		NCBI_TaxID=7227;		
RN		[1]		
RP		SEQUENCE FROM N.A.		
RC		STRAIN=ORESON-R;		
RA		Barrio R., Shea M.J., Carulli J., Lipkow K., Gaul U., Frommer G.,		
RA		Schuh R., Jackie H., Kafatos F.C.;		
RT		"The spalt-related gene of Drosophila melanogaster is a member of an		
RT		ancient gene family, defined by the adjacent, region-specific homeotic		
RT		gene spalt.";		
RL		Dev. Genes Evol. 206:315-325(1996).		
CC		-1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).		
DR	EMBL	Y07653; CABA68937.1; -		
DR	HSSP	P15822; 1B0.		
DR	FLYBase	FBgn0000287; salr.		
DR	InterPro	IPR003006; Iq_MHC.		
DR	InterPro	IPR000822; Znf.C2H2.		
DR	Pfam	PF00096; zf-C2H2; 8.		
DR	PRINTS	PRO0048; ZINCFINGER.		
DR	SMART	SM00355; Znf_C2H2; 8.		
DR	PROSITE	PS00290; Iq_MHC; UNKNOWN_1.		
DR	PROSITE	PS00028; ZINC_FINGER_C2H2_1; 8.		
DR	PROSITE	PS50157; ZINC_FINGER_C2H2_2; 7.		
KW		DNA-binding; Metal-binding; Nuclear protein; Zinc-finger.		
SO	SEQUENCE	1263 AA; 138991 MW; 2025B3BE67417622 CRC64;		
Query Match 14.0%; Score 742.5; DB 5; Length 1263;				
Best Local Similarity 23.5%; Pred. No. 7.8e-36;				
Matches 279; Conservative 143; Mismatches 379; Indels 385; Gaps				
QY	71	SNSSASAPRPEEHSRSQVMDTEHSNPPDSSSGCPDPPTWGERREDESSGGLVATGT	130	
DB	54	SNNGNTE-POMEAAVPE-SOTEREYAEERGEDEPE-----NSNEALDLSTSSGGR	104	
QY	131	AAGGGGGLI-----LASPKLGATPLPESTAPRPPPPPPPPGVSCHNLIPIL	181	
DB	105	ESLPGSGVHSLALQHTKVAVAGCAATACAGNHQA-----DLAMVQ	146	
QY	182	ELRLVQSQOIHOMQMTQICROY-----LLIG--SLGOTVGAPSPSELPTGAASSTK	234	
DB	147	STIFNVQROHLMQLOLQHLQSQLRRAEAAALGRSHSDDEEERPEPRPKQPTGKLR	206	
QY	235	PLPLPSPKPAQTGKTTASSSSSSSGAPPKQAFHLVPLPGSOHPPS-----VGG	288	
DB	207	EELELEQGESESDHREERNKTKDKRGTEDRKA-----EPGQYSMMCDLSSSLASS	260	
QY	289	VERSHKPTAPPS-----PALPGSTQQLTASFHLAEPGTTGLLAAGCLGAARGLAANAAPG	343	

147 STIFNVQRHLMQLQLIQLHLSQLKRAEAAALGRSHSDSEEEPEPEPKQPTNGLK 206
235 PLPLPSPKPAOTGKTTASSSSSSSGAEPKQAFPHLYHLSGHPSS-----VGG 288
207 EELEEQGESEDOESREENSKTDKRGTEBRKA-----EEGYQSMCDISSSIASS 260
289 VGRSHKPTAPS-----PALPGSDQLASPHLAPFGTGLLAOCGAGRLGAASPG 343
261 ITTNHDPAPAPNPNCLEMQRTBEVLDS-----ASQSIHAAMQOE----- 303
344 LKPKNGSELGYGEVVISLEKPGG--RHKCRCAKVFSGDSALQIHLRSHNGEPK 400
304 -----YSEYASKEAOSRGEIKFHCKYCGKIFGSYALQIHLRSHNGEPK 351
401 NVCGNFTTGNLKVHHRHREKPHVQMPHPVPHLDVITSSG-----LP 448
352 NVCGSKFTTKGNLKVHQRHTQILPFPMLPGVAPN-----VHSGGGYQGEQYPIRLP 406
449 YGMSV-----PEKAEAEAGTPGGGVERKPLVASTTALSATESITLSTGTS 495
407 FAPPAVPGOEONOVEEPEIRQEIPIVQAEDLSKPKVKE--KSHSPVEVKTPKE 464
496 TAVAPGLPFFNKFVLMKAVBPKSKADENTPPEGESGALAGVADSGSATRQOLS----- 548
465 VKTDAALPSESEK-----PEKEISK-----PVYTSRRNGSVKPKOTSAPSPQE 508
549 ----KLVTSLPSTALLNLKSTGSFPPVYLEPGA-----SPSTSKIQ 591
509 DRENDLVEHLIAKLVARSSASRESQPAEYSLAQMERITDKSWEDLEIDKTSSTSKIQ 568
592 IVEKIDROGAVAVASTAGAPTAPASSSASGPNOCVIGLRLVSCPRALRLHYOGHG 651
569 LVDNIE-----NKLTDPNQCFEQGVMSGRSSLOMIHIRTGTG 605
652 ERPPCKVCGRASFTKGNLRAHFVGHKTSPPARAONSCPIQCKFTNAVTLQOHVHML 710
606 ERPPCKICGRAPATKONLKAHMSIHKIKPPKRSQFKCPVCHOKFSNGIILQOHIRITM 665
711 ----GGQ-IPNGSALSEGGAQENSS-----EOSTAGPGSEFP 745
666 DDGSGGQGAAPANPGEERIGIEDONSNKSLGTSDTLDTSTTSDHSGGSESSOGGDFD 725
746 Q-----POSQPSD-EEMESEEEDEDEEEDVTDSDS-----LAGRSGE 784
726 EEMTMDSTDSRONNSAATATPHPLERERDREKERIRIPNDOSDERSHSNPDLTGGSE 785
785 SGGEKALISVAGDSE-----VSGAEE-- 806
786 SEMPMAMDLSPPSSNSGRIFATGLANGAAGGSGNGGLPMLGMPMPNLLMAAAREEM 845
807 -----VATSVAA----- 813
846 ALGHAAKAPPLPPLPGLFGMLHPPPNVCNLCFKMLPSTLAALESHQSEAKERATGHAQ 905
814 -----PTTVKEMDSNEKAPQHTLPPPP--PPDNLDPQ 845
906 RPOCSAGSPYGAKLTLNPNLFAKKPSSSSSGEKLPESSNPPPAENPPAPPIKEDP 965
846 PMEOGTSDVSGAMEEELKBCISSPMALT---QEGEGTSTPLVE-----ELNL 891
966 DQEQ-----LMVEEGASAGEGSGTGATSNYPQEGDAEQSLMKQMLAHRPPASPLDF 1018
892 PEA-MKKDGESS-----GRKACEVCGSPFTQOTALAEHOKTHPKXGPIFTVCRCRGF 944
1019 QOALMSAGSPPTSSLDPPVNNKHFCVCHCRNPNSSSSALQIHMRHTHTGDKP-FOCANVCQAF 1077
945 LDRATLKXHLMLAHNOVPPAPHPQONITATLSLVGCSSSIPSPGLS 991
1078 TTKGNLKVHM-GTHMMTNPTRSRGRBMSLELPMRGPNSGGQGHGSS 1123
9DGH3
ESULT 14

ID Q9DGH3 PRELIMINARY; PRT; 299 AA.
AC Q9DGH3:
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Putative spalt protein (Fragment).
GN SPAL.
OS Brachydanio rerio (Zebrafish) (Zebra danio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RA Camp E.M., Iardelli M.T.;
RT "PCR amplification of spalt exonic sequences from various
RT vertebrates."
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ293862; CAC05355.1; -.
DR InterPro; IPR000345; CytC_heme_bind.
DR InterPro; IPR000822; ZnF_C2H2.
DR Pfam; PF00096; ZnF_C2H2; 3.
DR SMART; SM00355; ZnF_C2H2; 4.
DR PROSITE; PS00190; CYTOCHROME_C; UNKNOWN_1.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 1.
DR PROSITE; PS0157; ZINC_FINGER_C2H2_2; 4.
KW DNA-binding; Metal-binding; zinc-finger.
FT NON_TER 1
FT NON_TER 299
FT NON_TER 299
SQ SEQUENCE 299 AA; 33064 MW; A9D03F130DA5FAB5 CRC64;

Query Match 12.5%; Score 663; DB 13; Length 299;
Best Local Similarity 41.7%; Pred. No. 7, 1e-32;
Matches 144; Conservative 49; Mismatches 66; Indels 86; Gaps 11;

QY 403 CGNRFTTGNLKVHHRHREKPHVQMPHPVPHLDVITSSGLPYGMSVPEKAEEA 462
Db 1 CGNRFTTGNLKVHHRHREKPHVQMPHPVPHLDVITSSGLPYGMSVPEKAEEA 462
QY 463 GTPGGGVERKPLVASTTALSATESITLSTGSTAVAPGLPFF-----NKFVLMKAVEPK 517
Db 59 ----SWLDSKPYLSTLTS-----SVGMLPPT---IPSLPFIKEENNSLAISPSHS 105
QY 518 SKADE--NTPPGSESGALAGVADSGSATRM-----QLSKLYTSL----- 554
Db 106 AKSDSGPADTPMKNTDS-----VLEEGESTLTPTSNGKAEENNOSSLLSMSSAVEGTIE 161
QY 555 -----PSWALLNLHLSKTSGFPPVYLEPLGASPSSTSKLQOLVEKIDROGAV 602
Db 162 YTTNSPPMATNPMLPMLMSQFKA--KFFEGGLDPL--QGSETSKLQOLVENIDRK--- 214
QY 603 AVASTASGAPITTSAPAPSSSASGPNOCVIGLRLVSCPRALRLHYOGHGERPPKCVGR 662
Db 215 -----VADPNECVICHRILSCQSALKMKNYRTHTGERPPKCVGR 254
QY 663 AFSTGNLRAHFVGHKTSPPARAONSCPIQCKFTNAVTLQOHV 707
Db 255 AFTTKGNLKVHMSVHRAMPRLVQHSQICQCKFTNAVTLQOHIR 299

RESULT 15
Q9DGH1
ID Q9DGH1 PRELIMINARY; PRT; 327 AA.
AC Q9DGH1:
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Putative spalt protein (Fragment).
GN SPAL3.
OS Brachydanio rerio (Zebrafish) (Zebra danio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.

Ox	NCBI_TaxId=7955;
RN	[1]
RP	SEQUENCE FROM N.A.,
RA	Camp E.M., Lardelli M.T.;
RT	"PCR amplification of spalt exonic sequences from various vertebrates."
RL	Submitted (Aug-2000) to the EMBL/Genbank/DBJ databases.
DR	EMBL: AJ293864; CAC05357.1; -
DR	InterPro: IPR000345; Cytc_heme_bind.
DR	InterPro: IPR008422; ZnF_C2H2.
DR	Pfam: PF0096; zf-C2H2; 3.
DR	SMART: SMO0355; Znf_C2H2; 4.
DR	PROSITE: PS00190; CYTOCHROME_C; UNKNOWN_1.
DR	PROSITE: PS00028; ZINC_FINGER_C2H2_1; 1.
DR	PROSITE: PS50157; ZINC_FINGER_C2H2_2; 4.
KW	DNA-binding; Metal-binding; Zinc-finger.
FT	NON_TER 1
FT	NON_TER 327
SO	SEQUENCE 327 AA; 35908 MW; 1623D9E63f7E9689 CRC64;
Query Match	12.3%; Score 654; DB 13; Length 327;
Best Local Similarity	40.8%; Pred. NO. 2.7e-31;
Matches 147; Conservative 47; Mismatches 78; Indels 88; Gaps 12;	
Oy	403 CGNRPTTGGNLVHFHRHREKYPHVQNMHPYVELLDVYTSSGLPYGMSVPPEKAEEA 462 : : : : : : : : : :
Db	1 CGNRSTGKNLKVHPGRHKKEKYHIQNNPYPVEYLVDNPVTSSGIPTYGMSLPKKPVTT- 59
Oy	463 GTPGGGEVKRPL--VASTTALSTATESLTLL-STGTSTAVA-----PGLPTF---- 505 :: : : : : : : : : : : : :
Db	60 ----WLDSKPILPTPTSVALOPLPPITPSIIGSDTSFSPLNSPQPAPSPSSECTS 114
Oy	506 ---NKFLMKAV-----EPKSKADENP-----CGSESATNGVMD-- 538 : : : : : : : : : :
Db	115 LSPNHLITETSTAQISSPPQPNLASNTIPPVLKPEALHLEPTNSTTRGETSISTASISQVI 174
Oy	539 -----SGSATRMQL-----SKIVTSLPSMALLTGNLTKSGSFPPVYLEPLGASPSETS 587 ::: : : : : : : : : :
Db	175 STTYITTTCTSTRQLDPVNSSSAVSHPSLSIQNQFK--KPFPGILDLM--QTSETS 230
Oy	588 KLOOLEVERKDROCAVAAVASTAGCAPTTAPARPSSASGPNOCVICLRVYLSCPARLRHYG 647 : : : : : : : : :
Db	231 KLOOLEVENIDKK-----MTDENOCVICHRLVSCQSALKMXYR 267
Oy	648 OHGGEPEPKCKVCGRAFSTRGNLRAHFPYGHTKTPPARAONSCICOKKFTNAVYLOOHVR 707 : : : : : : : : :
Db	268 IHTEHPERCKIGRAFITYGNLKTTHGVHRSKPLRVQHSCTPCQAKKFTNAVYLOOHIR 327
RESULT 16	
O9DGH2	
ID	PRELIMINARY; PRT; 288 AA.
AC	O9DGH2.
DC	01-MAR-2001 (TREMBLrel. 16, Created)
DT	01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT	01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE	Purative spalt protein (Fragment).
GN	SपालB.
OS	Brachydanio rerio (zebrafish) (zebra danio).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC	Cyprinidae; Danio.
OX	NCBI_TaxId=7955;
RN	[1]
RP	SEQUENCE FROM N.A.
RA	Camp E.M., Lardelli M.T.;
RT	"PCR amplification of spalt exonic sequences from various vertebrates."
RL	Submitted (Aug-2000) to the EMBL/Genbank/DBJ databases.
DR	EMBL: AJ293863; CAC05356.1; -
DR	InterPro: IPR000345; Cytc_heme_bind.
DR	InterPro: IPR008422; Znf_C2H2.
DR	Pfam: PF0096; zf-C2H2; 2.

Query Match	Best Local Similarity	Score	DB 13	Length	288
Matches 142	Conservative 39	Mismatches 69	Indels 91	Gaps 9	
Query	403	CGNRFTTRGNLKHFEHREKYPHYOMNPHPEHLDYVITSGLPYGMVPEPKAE	462	1	CGNRFTTRGNLKHFEHREKYPHYOMNPHPEHLDYVITSGLPYGMVPEPKAE
Db	1	CGNRFTTRGNLKHFEHREKYPHYOMNPHPEHLDYVITSGLPYGMVPEPKAE	55		
Query	463	GPFGGVERKPLVASTALATESITLLSTGTSTAVAPGLPFNFKVLMKAVEPKSKADE	522		
Db	56	PLSLMDSKPLVGNISIGFMLSSD	98		
Query	523	NRPPEGE	549		
Db	99	HSEVVELTRKINGHQBGFVCSPLIISNEKEQEVNORLSVSTLRSRESREDIAINTS	158		
Query	550	LTSLPSWMLLTJNHLKSTGSPFPFYVLEPLGASPSETSLQOLVKEIDRGAVAVA	605		
Db	159	VNTGLTIITKLKSEQLEAKFL-LGSLP	203		
Query	606	STASGAPPTTAPAPSSSSASGPNOCVLCRLVISCPRALRLHYGQHGERPKCKVCAAFS	665		
Db	204	YTDNEGICHRVLSQSLRHMFRHTGERPKRCVCGRAFT	246		
Query	666	TRGNLRAHFVGHKTSPPARAONSCPTCKKFTFNATYLLQOHV	706		
Db	247	TKGNLTKHYHSIHRSMPLRLIIONSCPICOEKFPMANVLQOHI	287		
RESULT 17					
ID	09PVNS	PRELIMINARY	PRT	549	AA
AC	09PVNS				
DT	01-MAY-2000	(TREMBLrel. 13, Created)			
DT	01-MAY-2000	(TREMBLrel. 13, Last sequence update)			
DT	01-JUN-2002	(TREMBLrel. 21, Last annotation update)			
DE	Xsal-3', long form (Fragment).				
GN	Xsal-3',				
OS	Xenopus laevis (African clawed frog).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;				
OC	Xenopodidae; Xenopus.				
OX	NCBI_TaxID=8355;				
RN	[1]				
RA	SEQUENCE FROM N.A.				
RA	MEDLINE=99458636; PubMed=10527856;				
RA	Onuma Y., Nishinakamura R., Takahashi S., Yokota T., Asashima M.,				
RT	"Molecular cloning of a novel Xenopus spalt gene (Xsal-3)." ;				
RT	Biochem. Biophys. Res. Commun. 264:151-156(1999).				
DR	EMBL; AB030828; BAA85902.1; -.				
DR	HSSP; P15822; 1BBO.				
DR	InterPro; IPR000822; znf_C2H2.				
DR	Pfam; PF00096; zf_C2H2_5.				
DR	ProDom; PD000003; znf_C2H2_1.				
DR	SMART; SM00355; znf_C2H2_5.				
DR	PROSITE; PS00028; ZINC_FINGER_C2H2_1; 5.				
DR	PROSITE; PS0157; ZINC_FINGER_C2H2_2; 5.				
KW	DNA-binding; Metal-binding; Zinc-finger.				
FT	NON_TER				
FT	NON_TER				
SO	SEQUENCE	288	AA;	32261	MM; A07A1FBDEAF42 CRC64;

[illegible]

Best Local Similarity 22.1%: Pred. No. 1.9e-14; Matches 162; Conservative 92; Mismatches 280; Indels 271; Gaps 29;

QY	22	GGCASEEHHNOVCAKCAQAFSDPTPEFLAHQNSCTCDPRVMVILGGQENPSNSASSAPRP	81
Db	51	GGALPLEDRSNILPHFSVPFANPOFL-----SLCAQ-----LGNSSNRVSSSTAS-----	96
QY	82	EGHSRSQVMDTEHSNPDSDGS---SGRPDPVWGPERRGSESSGQFLVAATGTAAAGGGGL	138
Db	97	-----TTSCTPIQSCQSQSFSSPALITHVLDADHDEQEIFSCDVCCTTFFSNGODI	146
QY	139	-----ILASPKIGATPLPESTPAPRRPPRRPPRRPGVSGSHLNLPLLELRVLDOQR	191
Db	147	REHKQKTLAS---RSTSVSPSTIPSSVCFLSTPTTPCL-QPSINSEIGSTELREDEDEE	202
QY	192	IHOMQMEJOICQVILLGSLGQTFGAPASPSLEPCTGAASSTKPLPLPSPIKPAQCGKT	251
Db	203	DMVEDEGEHVAHQ--LGHLLQ-----KS	224
QY	252	TASSSSSSSSSGABPPKQAFPHLYHPLIGSQHDFSVGVGRSHKPRPAPSPALGSTDOLI	311
Db	225	DDKSKMASTLNNHAPFPFAAFPMPPFIMRQFDF-----PRADVFA	265
QY	312	ASHPLAPPGTTGLLAQCLGAARGLLEAASPELLPKPKNSGSELGYEYLSLEKPGGRHK	371
Db	266	AGRHDDDDDEALMEISTSDAEKIRALV-----GDKAVPTTDP--NQ	306
QY	372	CRCFCAKVFSGSDALQILHRSHTGERPYKCNVCGNFFTTTGNLKVPHRRREKYPHYQOMP	431
Db	307	CILCRVRLSCKSALQMIYRTTGERPFRCIKQICRAFTTKGNLKTGHVGRSHK-----	359
QY	432	HPVPEHLDYVITSSGLPEYMSVPEKA-----EEEAGTGGGV--	469
Db	360	-----SFRGLP--ISLPPOLAMHQHQIAPQRIHINPPTSASASAAVAQ	406
QY	470	-----EKRPVASTALSTATE---SLT-----LSTGSTANA--PGLPTF-----	505
Db	407	IQASOOCPIQOARLNGELAVHTEHNSLSLQPRVMPPTPTTRQTPPEPFVFTTPPS	466
QY	506	-----NKFVLMKAVEPKSKADENTPGSGSALAGVADSGSATPMQSLKSLTSPMSA	558
Db	467	LNATDMSQTNLANILLSAQLKNDSS--PNTDTSSV-----EKKTRDDPPKMASLSDS--	517
QY	559	LITNLHKSTGSPFPFVYLEPLGASPSSETSKLOOLYER--IDROGAVA-----VAST	607
Db	518	-----NSDSSSS-----VRDILSESEPEFKLKELEPPILEQOVSTTPMPKNENPPLAMQ	569
QY	608	ASGAPITTSAPAPSSSAGPNOCVICLRLSCPRALRLYHGGGERPFRCCKVCGRAFSTR	667
Db	570	KWMAEEPPPPROMPYVLSKQOCGVCFKHFSSSSSALQIMHRTGTDKPFCDCMGCAFTRR	629
QY	668	GNLRATFVGH--KTSPARA-----	685
Db	630	GNLKYMHGTHSMQOSPBRGRIRIDVASSVTEKPMLOSLILPTSGAPGASPLAMLCPNGL	689
QY	686	-----QNSCPICOKKFTNAVTTLQOAHYRMHLGGQIPNGSGA	720
Db	690	SGLEMMAMLMRLTYGCVQCQKQOSPNELEQHLKEHNL-----NOSA	730
RESULT	19		
Q61479	061479	PRELIMINARY;	PRT; 2282 AA.
AC	061479:		
DT	01-NOV-1996	(TREMblrel. 01, Created)	
DT	01-NOV-1996	(TREMblrel. 01, last sequence update)	
DT	01-MAR-2002	(TREMblrel. 20, last annotation update)	
DE	DNA binding protein Rc.		
GN	KRC OR Rc.		
OS	Mus musculus (Mouse).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
OX	NCBI Taxid=10900;		

RX MEDLINE=97349984; PubMed=9205841;
 RA Nagase T., Ishikawa K., Nakajima D., Ohira M., Seki N., Miyajima N.,
 RA Tanaka A., Kotani H., Nomura N., Ohara O.;
 RT "Prediction of the coding sequences of unidentified human genes. VII.
 RT The complete sequences of 100 new cDNA clones from brain which can
 RL code for large proteins in vitro.";
 RL DNA Res. 4:141-150(1997).
 CC -I- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
 DR EMBL: AB002324; AAC20784.1; -
 DR HSSP: P08047; 1SP2;
 DR InterPro: IPR000822; Znf_C2H2.
 DR Pfam: PF00096; zf-C2H2; 19.
 DR PRINTS: PR00048; ZINCINGER.
 DR SMART: SM00355; ZNF_C2H2; 19.
 DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; 19.
 DR PROSITE: PS00157; ZINC_FINGER_C2H2_2; 19.
 KW DNA-binding; Metal-binding; Nuclear protein; zinc-finger.
 FT ¹_{NON_TER}
 SQ SEQUENCE 927 AA; 102756 MW; 176C78A9C407F8C0 CRC64;

45;

QY	67	GNPNSSSSXSSAPR-----PEG-HSRQOMDTSHNPDSG--SSG--PRDPTWGE	113
Db	10	EXDPRVLGAQSVPRGRALKGLSPFLGDSARFLRPDPRAGPNNYAVLSSGMETPLMGFD	69
QY	114	RGESSGQPLVAATGTAA-----GGGGILIASP-----KIGATPLPRESTAP	158
Db	70	LOGPEOSPN--DAHRAESENEEESPPOESSGEETIIMGDPAQESKSTESTERSQD	127
QY	159	PPPPPPPPGVSGGLNPLILEELRVLQDQRIHQOMTPOICQIVLLLSLGTVCAP	218
Db	128	PSVQONPPTPLGHSN-PL-----DHQILPLD-----P	152
QY	219	ASPSSELP-----GTGAASSTKPLPLTSP--IKPAQGTMTTASSSSSSSS	262
Db	153	PAPVAVPTPSDWTKACEASQWQMGALITWNNSPVVPAPEPSIRELVQR	200
QY	263	GAEPRKQAFHLLHYRGSQHPREYGVGVRSHKTPRAPSPALPGSTDQLIASPLAFGTT	322
Db	201	-----PAGAKRPYCNBEGKS-----FSQMS	221
QY	323	GLLAOCSGANGLEAAAPRGLIKPKNSSSELGEYVITSLLEKPGCRHKCFKCAVFGSD	382
Db	222	KLIRHQRIHTGERPNMCSBG--KSTQSSHLVQHQHTHGEKP--YKPCDGCFCFWS	276
QY	383	SALQIHLHSHGERPYKNCVCGNRFITRGMLKXHFHNRKYP-----	425
Db	277	SNLVQHQRIHTGEKPYKCECEKAFQSTNLIKHORSHTGEKPYKCECRAPFRSSDLI	336
QY	426	-----HYOMNPRPEHLDLYITSSGLPYGMSVPRPEAE--EATPGGVGRKPLVASTT	479
Db	337	QHQATHGEKPYKCE-----CGKRRGQNNILKQKQIHAG-----EKRYRCEC	381
QY	480	ALASATESLTLIS--TGTSTAVAP-----GLPTFNKVLKMAVEPKSKADENTPRGSECSA	532
Db	382	GKSFQSSSELQHQRIHTGEKPYECLECKSPFQSSSTLIKQRIHLREDFKCP-----	435
QY	533	IAGVADSSSATRMOJLKLVTSLPSMALLNNHLKSTSGSPRPYULPLGAPSEFSKL--Q	590
Db	436	VCGKTFITLSAT-----LLHQRIHTE--RPYKCPCEGKSEFSVSSNNLHN	478
QY	591	QVLEKIDROGAVAVASTASGAEPTTAPAPADSSASGPNQVICTLRVILSCPRALRLHYGQNG	650
Db	479	QRIIRGER-----PYICADCKSFIMSTILRHQRIHT	511
QY	651	GERPFKCVGGRAPSTRGNLRAHFVGHKTSAPARAQNSCPICQKKTNNAVTLOQHVAMHL	710
Db	512	GEKPYKCDCKSFIR-----SSHLLQRIKTHTEKPYKCECKSFQSSSNLLTHVETHM	567
QY	711	GGQI-----PENGSAISEGGAAQENSDEOSTASGSGSFPOQSOQSPPEEMSEEEDEE	767

568 DENLFVCSDCSKAFLEAHLEHLEQHRVIERH-----GKTPARRAQ----- 605
768 EEEEDVTEDESLAGRSES-----GGEK-----AISVRGDSSEVSGAEEEVATSYAAPTYK 818
606 -----GDSLLGLGDDPSLLTPRRPGAKPHKGLCVKGKGFNDE--GIFMORHITIGENPYK 656
819 EMDSENEKAPQHTLPPRRPPRDLDHPOMEGOTSDVSGAMEEAEAL-----EG1 867
657 NAGG---LMAAARPR-----QLRSRPLPRGRNS-YFGAAGRAAPROPPLKPRPEGEGF 708
868 SSPMAALTOE-----GE-----GSTPLV---BELNLPKAMKDPGES 902
709 SORRGLLSKTYICSHCESFLDRSVLLHQHLTHGNKRPFLFPDYRIGLEGAGSPPLS 768
903 SGRKACEVCGSFPQTALAEHOKH-----PKDG-----PL---FTC 937
769 GKRPKCEPKQSFGSLSELLHOKYHAGKSSQKSPELGSSVLEHLRSLPLGARPYRC 828
938 VFCRQGLDRATLTKKHMLLAHNOVPPFAPHPGONITATLSLVRGSSSTPSP 988
829 SDCRASFLDRVALTRHGOETHNQEKPRNPEDPRPEAVTLSTDQEGEGTPTP 879

RESULT 22
ID 015090 PRELIMINARY; PRT: 1300 AA.

015090: 01-JAN-1998 (TRMBLrel. 05, Created)
01-JAN-1998 (TRMBLrel. 05, Last sequence update)
01-JUN-2002 (TRMBLrel. 21, Last annotation update)
KIAA0390 protein.
KIAA0390.
Homo sapiens (Human).
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
NCBI_TaxID=9606;
[1]
SEQUENCE FROM N.A.
TISSUE-BRAIN:
MEDLINE=97349984; Pubmed=9205841;
Nagase T., Ishikawa K., Nakajima D., Ohira M., Seki N., Miyajima N.,
Tanaka A., Kotani H., Nomura N., Ohara O.,
Prediction of the coding sequences of unidentified human genes. VII.
The complete sequences of 100 new cDNA clones from brain which can
code for large proteins in vitro.";
DNA Res. 4:141-150(1997).
-1 SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
EMBL: AB002388; BAA20844.1; -.
HSSP: P15822; 1BBO.
InterPro: IPR000822; ZnF_C2H2.
Pfam: PF00096; zf-C2H2; 10.
PRINTS: PR00048; ZINC_FINGER.
PRODOM: PD000003; ZnF_C2H2; 1.
SMART: SM00355; ZnF_C2H2; 10.
PROSITE: PS00028; ZINC_FINGER_C2H2_1; 6.
PROSITE: PS0157; ZINC_FINGER_C2H2_2; 8.
DNA-binding: Metal-binding; Nuclear protein; zinc-finger.
SEQUENCE 1300 AA: 141416 MW: 85855B7047653664 CRC64;

Query Match 6.3%; Score 336.5; DB 4; Length 1300;
Best local similarity 20.8%; Pred. No. 1e-11;
Matches 193; Conservative 102; Mismatches 317; Indels 317; Gaps 37;

258 SSSSSGAEPRQAFPHLHNP-LGSHPFPSVGVGRSK-----PTPAPSPAL 303
11 SSAFEPRP-----HLSGPVNGOYAMS---QKLRQITSQLSHAPPELHPRNPDEKRP 60
304 PGSTDOIAPSHLAFR-GTTGLAOCIGAARGLEAASPLLPKNGSGELGYGVVIS 362
61 PASLEEKAHYVMSGQPMGSMALLANQLEGREVDTSLNGRDVLDQFLNQ-NIGINSQMSD 119
363 LE---KPGGRHKRCARCAVFGSDALSALQILHLSHTGERPYKCNVCGNRPETTRGNLKVHFR 419

Db 120 IEDDARKNRKTPRCPCGKRRFRNSILSMKRTHTGKEPRFCPCYCDHRAQOKNLKTHLT 179
Qy 420 H-----REKYPHVQNPHPVPEHL---DYVITSSGL-PYGSVPRPEKAEAGTPG 467
Db 180 HKLGNLGKGRGRVRENRLHLEERAILRDKOLKSLDRPRDLKPRHNAO----- 232
Qy 468 GVEKPLVASTALSTATESLTLSTSTSTAVAPGLPFPKPFVLMKAVPKSKADENTPRG 527
Db 233 ---APLAACLTALQANHSV-----PDVAHPVPS-----PK-----PAS 262
Qy 528 SEGSAIAGVAD-----GSATRMOLSKLVTSLPSMALLTNHLKSTGSFPFVYLEPLGA 581
Db 263 VQEDAVAPAAFRCTGCKGFKKRE-----ELDRHITLNR-PYCTTCLDFPA 309
Qy 582 SPSETSKLQOLVEKIDROGAVAVASTAGAPPTSAPAPSSASGPNQCVICLRLVLSCPRA 641
Db 310 SOBE-----ELISHVEK-----AHITSAOGGQPN----- 335
Qy 642 LRLHYGQNGER---PFCKYCGRAFTSGNLRAHPYGHKTSAPARAQNSCPICQKFTN 698
Db 336 -----GGGQSANEFRCCEVCGVFSQAMFLKGHMKRHKDS---FHHCOICGRFKE 384
Qy 699 AVTLQOHVRMILG-----GQIPNGSALSSEGGGAOE----- 730
Db 385 PWFLKNMKVHLNKLSTYKNSPSPDEVPYPMGMSQZAHANLYSRYLSCLSGFMTPKXA 444
Qy 731 NSSEOSTASGSPFPQSOQ---PSPEEMSEEEDEE----- 767
Db 445 GLSEPSOLYGKGLRPMKEKALGLKLSPISSMAHGVPEGDKHSLGLCLNVLPLKSSCIE 504
Qy 768 -----EEDVDYD-----EDSLAGSGSEGGKA 790
Db 505 RLQAAKAAEKNDPVNSYQAWOLMARGAMEHGLSKENHPLQRNHEDTLNAGVLFDEKRR 564
Qy 791 --ISVRGD-SEEVSGAEAEVATVAA---PTTVKEMDS-----NEKAQ 828
Db 565 EYLVGADGSKQKMPADLVHSTKVSQGRDLPSKLDPLESSRDFLSHGILNQTLEYMLQCG 624
Qy 829 HTLPPPPPPD-----NID-----HPQME 848
Db 625 NMKEKPTCEPCGRFRTYHQVYVHSRVHKKDRKGEEDGLHVGLDERRGSGSDQESQSVS 684
Qy 849 QGSDVSGAMEEAKLEGISSPMALTOGEGSTPLVEELN-----LPKAMKDP 899
Db 685 KSTTPGSSNVTSESGVGGSLGTSQAQEDSPHPSSESSDIGEAGRSAGVOQRLPKR 744
Qy 900 GESSGRKACEVCGSFPQTALAEHOKTHPKDGPLFTCFRCRQGLDRATLTKKHMILLAH 959
Db 745 SLGSAMKDCPYCGKTFRTSHNLKVNHRITHTGEK-YKCPHCDYAGTQSASLKYH-LERNH 802
Qy 960 Q-----VPPFAPHPONITATLSL 977
Db 803 RERONGAGPLSGOPPNODHKDEMSSKASL 831
RESULT 23
Q96K58 PRELIMINARY; PRT: 619 AA.
AC Q96K58;
DT 01-DEC-2001 (TRMBLrel. 19, Created)
DT 01-DEC-2001 (TRMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TRMBLrel. 21, Last annotation update)
DE CDNA FLJ14492 f1s, clone MAMMA1002937, weakly similar to zinc finger protein 135.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=MAMMARY GLAND;
RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,

Db 381 -----SHLOYHLRLHTGERPYRCPDCP-----KA 404
QY 521 DENTPPSEGSALAGVADSGSATFRMOLSKLVTLSPSMALLTNHLKSTGSPFPVYLEPLG 580
Db 405 FKNF-----SCLGPH-----414
QY 581 ASPESTSKLOOLVEKIDROGNAVAVASTAGAPTTSAFAPSSASAGNOCVICLRVLSGPR 640
Db 415 -----RDL-----HTGERPHACPTICGKAFOTGS 437
QY 641 ALRLHYGNGGERPFCVKYCGRAFSTRGNLRAHFVGHKTSAPARAONSCPTCQKFTTNAV 700
Db 438 NLRHQRHRTHTGERPACSHCGKTFTHSSNQLH---QRTSHSAR-PHCQPLCPKAFVMS 493
QY 701 TLQGHVRMHLGQIPINGSALSBCGGAOENSSQSTASGSPPOPOSQOPSPREEMSE 760
Db 494 YLQHLRLFTN-----AAGFKGSGRPP-----512
QY 761 EEEDEEEEDVTDSDSLAGRSGESGGEKALISVNGDSEEVSGAEDEVATSVAPTTVKEM 820
Db 513 -----ALTPQRDGPVLQAA---LSLEVTAP-----534
QY 821 DSNF-----KAPD-HTLPPPPPPDNL-----DHQPMEGTSDVSGAMEEBAKLEGIS 868
Db 535 DAHFFLLQTPQGLQIPSPBPAPQKILILPTAPQPPKHGESPPTPGOSLLLVSTGTT 594
QY 869 SP---MAALTOGEGCTPLVEELNLPEAMKKDRGESSGRACE 909
Db 595 LPTURLAQAVTAVPOGTGTLVLQGLPEQPLRPAIGIRHQAAYE 638
RESULT 25
Q9P2A7 ID Q9P2A7 PRELIMINARY; PRT; 1258 AA.
AC Q9P2A7;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE KIAA1441 protein (Fragment).
GN KIAA1441.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-BRAIN;
RA MEDLINE=20181126; PubMed=10718198;
RA Nagase T., Kikuno R., Ishikawa K., Hirosewa M., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XVI.
RT The complete sequences of 150 new cDNA clones from brain which code
RT for large proteins in vitro.";
RL DNA Res. 7:65-73(2000).
DR EMBL; AB037862; BAA92679.1; .
DR HSSP; P17789; 2DRP.
DR InterPro; IPR000345; Cytc_heme_bind.
DR InterPro; IPR000822; Znf_C2H2.
DR Pfam; PF00096; zf-C2H2; 12.
DR SMART; SMD0355; Znf_C2H2; 12.
DR PROSITE; PS00190; CYTOCHROME_C; UNKNOWN_1.
DR PROSITE; PS00128; ZINC_FINGER_C2H2_1; 9.
DR PROSITE; PS50157; ZINC_FINGER_C2H2_2; 8.
DR DNA-binding; Metal-binding; zinc-finger.
FT NON_TER 1
FT SEQUENCE 1258 AA; 131684 MW; 04D3F1BELD497782 CRC64;
Query Match 6.1%; Score 325.5; DB 4; Length 1258;
Best Local Similarity 20.4%; Pred. No. 4.3e-11;
Matches 245; Conservative 126; Mismatches 445; Indels 385; Gaps 54;
QY 12 GCGCGEPAERGGDASEE-----HHPOVCAKCAAOFSDPTEF-----LAHQNSCCTDPPV 60
Db 59 GGP-GRPEPGVSESEDTAAASAGDGPVPAQASDHGLRPPDISVSVIVKNTVCPDQSE 117

QY 61 MYIIGQENPNSNSASAPRPEGHSSQ-----VMDTEHNPDDSGSPDPPTWG 111
Db 118 ALAGSAGDAQADAVYTKGKRGVPHRMQNGFSPSPSLPTGRPHSAPRSGS-----TW- 170
QY 112 PERGEESSGOFLVAATGTAAAGGGGLILASPKLGAPLP-----PE-STAPPPP 161
Db 171 -KEKME-----GKPRLDLFAHGRPERGDHSDLPESA 202
QY 162 PPP-----PPP-----PGVSGHLNIPLLELRVYLQKQI--HOMQMTQ 200
Db 203 PSPREGALTPPPFPSSFELAOENGP-----MOPVSSPPLGALKQSCSPHNPQVLQ 257
QY 201 ICRQVLLSLGQTVAGAPSPSELPTGA-----ASTKPLDLPFSPRIKPAQTGK 250
Db 258 QGS-----GSSFKATDIPASAPPPVAVGPFVKQSPGHQSLPASKVYVQCQLKEEDDE 312
QY 251 TTAASSS-----SSSSSGAE-----PRQAFPHLYPLHSQHPFSV-----GGV 289
Db 313 GPVDRKSSPSPQSPSSGAELAEDSDNDSPASS-----SSRPLKVRIKTIKTSQNI 363
QY 290 GR-----SHKPPAP-----SPALPGSTDQLASPHL-----316
Db 364 TPTVTQVPSPDPPAPLAEGAFILAASLLKLSPATPTSEGRKVVSVQLGDGTRLKGTLP 423
QY 317 -----APRGTTL-----LAAQILGARGLEAASPG 343
Db 424 VATTIONASTRAMLAASVARKAAVPLRPGGATSPKMLAKNVGLVPRQALPRADGAGLGTTG 483
QY 344 -----LKR-KNGSG-ELGYEYIS-----361
Db 484 OKYNGASVWVQPSKATPTSGTGGTIVSRQTSLVEAFNKLINSKLLPAYRPNLSPPA 543
QY 362 --SLEKPGGHRKCFKAKVNGSDSALOHLNHSHTGERPKKCVNGGR--FTTGNIKLVH 417
Db 544 EAGLALRPTGYRCLGCDADVLEKSLARHYRRSRRIEYTCNCAIRLYEFNCSLLHA 603
QY 418 HHRERK-----YPHVQNNPVRPHLDVITSSGLPYGMSVPEKAEEBAGTPGGVERK 472
Db 604 REHKDKGLVWCQSHLVNRPALDQWVGORDITPLRP--VAVPR--VSGPLALPALGEG 659
QY 473 PLVASTTALSATESLTL-LSTGTSTAVAPGLTFENKEVYLMKAVEPKSKADENTPPSEGS 531
Db 660 AITSSAITTVAAEAPVPLST-----EPRAAPATSAUTCFCRLECKECCGDK-----706
QY 532 ATAGVA-----DSSATPMOLSKLVTLSPSMALLTNHLKSTGSPFPVYLEPLGASP 583
Db 707 --AGMAAHFOQLGPRPAGATSNVCPCTPMILPNKCSFSNQRNKNRP-PHVCPECGG--761
QY 584 SETSKLOOLVEKIDROGNAVAVASTAGAPTTSAFAPSSASAGNOCVICLRVLSGPRALR 643
Db 762 ---NPLQANQTNHREACLNHSRVG-----YRPPSCGVVGGVNSIK 801
QY 644 LHYGNGGERPFCVKYCGRAFSTRGNLRAHFVGHKTSAPARAONS-----CPTCQKFTN 698
Db 802 SHIQTSHCEVFNKCRIPMAFKGSPSAHANL--YSQHFPFQOALILYICAMCDTFTTH 859
QY 699 AVTLQGHVRMHLGQIPINGSALSBCGGAOENSSQSTASGSPPOPOSQOPSPREEM 758
Db 860 KPLLSHFDOHL--LP-----ORVSYFKRPSCLLRAQKTYML-----EHL 898
QY 759 SEEEDEEEEDVTDSDSLAGRSGESGGEKALISVNGDSEEVSGAEDEVATSVAPTTVK 818
Db 899 KNTHQGRLEF-----TAGKGA---GGALLTPKTEBELA-----VSGGGAAPATEE 942
QY 819 EMDNSNEKAPQHTLPPPPPPDNLHPQPMEOGTSDVSGAMEEBAKLEGISPMALALQEG 878
Db 943 SSSSE---EEEVSPSPPRPAKRPR--RELGSKGLKGG-----GGG 980
QY 879 EGTSTPLVEELNLPEA-----MKKDPGSSGGRKACVCGSPPTQATLEHOK--TRPKD 931
Db 981 PGGWTCGLCHSWPBRDEYVAHMKREHGKSVKKFPCRLCERSFCAPSLRRHVAVNHNGI 1040

Drosophila

932 GPLEFCVCFRCROG---FLDRATLTKKMLLAHHQVPFPAHGRQNINATLSLVPCSSSIIPSP 988
| | | | | | | | | | | | | | | | | | | | | |
1041 KRVYCRCTCEKRFTTSSLRLTEKHVRHGEL--QLGAOSPEGRITLLA--RSSSARAAGP 1096
y 989 G 989 y
b 1097 G 1097 b

ESTUT 26
D . Q24219 PRELIMINARY; PRT; 744 AA.
C Q24219;
G 01-NOV-1996 (TREMBLrel. 01, Created)
C 01-NOV-1996 (TREMBLrel. 01, last sequence update)
T 01-JUN-2002 (TREMBLrel. 21, last annotation update)
E BOMEL protein.
N BOWI OR CG10021.
S Drosophila melanogaster (fruit fly).
C Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
C Pelecygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
C Ephedroidea; Drosophilidae; Drosophila.
NCBI_TaxID=7227;
[1]
P SEQUENCE FROM N.A.
X MEDLINE=96272178; PubMed=8670819;
Wang L., Coulter D.E.;
T "Bowel", an odd-skipped homolog, functions in the terminal pathway
during Drosophila embryogenesis.";
EMBO J. 15:3182-3196(1996).
[2]
P SEQUENCE FROM N.A.
X MEDLINE=97032935; PubMed=8878683;
Hart M.C., Wang L., Coulter D.B.;
T "Comparison of the structure and expression of odd-skipped and two
related genes that encode a new family of zinc finger proteins in
drosophila".
Genetics 144:171-182(1996).
[3]
P SEQUENCE FROM N.A.
X STRAIN-BERKELEY;
C Celniker S.E., Aghavani A., Arcania T.T., Baxter E., Blazes R.G.,
Chew M., Doyle C.M., Farfan D.E., Flanagan J., Houston K.A.,
Hummel S.L., Karra K., Kearney L., Kim S.H., Lee B., Lomocan M.A.,
Mazda P., Mok M.S., Mostreli A.R., Mostreli M., Nixon K., Pacleb J.M.,
Park S., Pfeiffer B., Punch D., Shih E., Twomey B., Wan K.H.,
Whiteley K.R., Yee A., Zhang R., Zierak L.J., Kimmel B.;
Submitted (JUN-1998) to the EMBL/GenBank/DDO databases.
L - SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
EMBL: U58282; AB017949.1; -.
R EMBL: AC005149; -. NOT_ANNOTATED_CDS.
R HSP: P07248; ZADR.
R Flybase; Fggn0004893; bowl.
R InterPro: IPR000822; ZnF_C2H2.
R Pfam: PF00096; zf-C2H2. 5.
R PRINTS: PR00048; ZINC_FINGER.
R ProDom: PD000003; znf_C2H2. 2.
R SMART: SM00355; znf_C2H2. 5.
R PROSITE: PS00028; ZINC_FINGER_C2H2_1. 5.
R PROSITE: PS50157; ZINC_FINGER_C2H2_2. 5.
DNA-binding; Metal-binding; Nuclear protein; Zinc-finger.
CONFLICT 196 196 S -> F (IN REF. 3).
CONFICT 634 634 L -> P (IN REF. 3).
CONFLICT 720 720 P -> A (IN REF. 3).
SEQUENCE 744 AA; 79813 MW; ED565ZCE6FCDD9684 CRC64;

Query Match 6.1%; Score 322.5; DB 5; Length 744;
Best Local Similarity 18.3%; Pred. No. 3.4e-11;
Matches 195; Conservative 90; Mismatches 281; Indels 501; Gaps 37;

90 MDHEHSPPDGSGCPDPPTMGPERGERGESSQFLVA-ATGTAAAGGGGLASPKLG-- 146
1 MPRESSSETISGGGGGAIPMRPSR-----MDGFNNSMAAAAAVCGGPGAADRNIGS 55

QY	147	-----	ATPLPESTP-----	APPEPP-----	164
Db	56	GGSGGGGONGNGDSRNSASAKI	SAVETOLATOIHLA	GLHGPPPPPSHHRETSAPVYL	115
QY	165	---PPPGVSGHLNIPLELRYL	QOIOHOMQTEQICR	OVLLGSLGTVGAPASP	221
Db	116	PTGVRRGSGNSNVEI	IAMADRKELALRE-----	AAAAAMLGRRPGGCGP	163
QY	222	SELPGTAASBTKLPL	FPSPIKRQOTKTTA	SSSSSSSSGAEPPKQAF	PHLYHPLGSO 281
Db	164	G-VPPPV-----	LYGP-----	AGVPP-----	180
QY	282	HPFVSQVGRSHKPT	PAPSPALPGSTDOLIAS	PHLAFPGTGTGLLAQO	CGAARGLFMAAS 341
Db	181	-PLTG-----	PGSPGTAGS-----	SP-----	FRPGAAAAALFPFGLGCGMAAGL 221
QY	342	PLLK -PKNSGELGYE	EVISLEKPGGRHKCR	FCACAFVGSALQIHL	RSHTGERPYKC 400
Db	222	RRLRAFGRAS-----	---RPKQFOTCKFC	PNQPTKSYNLLIBRT	HTDERPYSC 268
QY	401	NVCGNRTTGNLKV	HHRHREKTPHYOM	NHPRVPHLDYVIT	SSGLPYGMSVPREKAE 460
Db	269	DTCGAARRDHLRDH-----	RYIH-----		288
QY	461	EAGTPGGVVERKPL	VASTALSTATESITL	STGSTAVARGLPT	NKFLVLMKAVEBKSA 520
Db	289	-----			288
QY	521	DENTPPGSEGSAL	AGVADSGSATRMQSL	KLVTSLSMALLTNH	LKSTGSPFPYVLEPG 580
Db	289	-----			288
QY	581	ASPSETSKLOQLEK	IDROGAVANA	ASTASAPRTSAPAR	SSSAGPNCVILRVLSGR 640
Db	289	-----			307
QY	641	ALRLYHGOHGER	PFCKVCGRFSTR	GNILRAHFNHKTSP	AAARONSCPIQCKFTNAV 700
Db	308	TALVHKLHMEESP	HKRCVCSRSRPNQ	SNLKHLLTH-----	TDHKRYESSCGYVRRAC 363
QY	701	TLOQYRMLHGOI	PNCGSA-----		LSBGGGAA 728
Db	364	DIRRALTLHA	VGUNSGDYVDGEBE	DARMLSGDEBESL	LEVDSPROSPVHMLGEGSG 423
QY	729	OENSSEOSTAS	CGSPFOPOSOQSP	PEEMSEEBEDE-----	BEEDYUTDED 776
Db	424	EKSESERMRLK	RAALDHES-----	EEEFDDFEEELQD	LPYHDLPREDDDFRDED 478
QY	777	-----	SLAGRSES-----		GGBK -A1SVKGDSEEVSG 802
Db	479	EOQAEVALVAR	FOQASKAATASQSS	SVGTRKEROGVTH	CHNEGGETYMRPGEKHQEP 538
QY	803	ABEEVAT	STAAPTIVKEMDS	NEKARQUTLPP-----	PPDNLDPH----- 844
Db	539	GNSGIA	SLPVPBFVR-----	YSVPPGAAGPP	APRPARPPTQOHDPHILPRN 588
QY	845	-----	QMEQOTSPVSG	AMEEAEKLEGI-----	SSPMALTOGEGSTPLVEEL 889
Db	589	GDPIRLIL	VARROLHHRKSL	LSKA-----	GVPPPHPTITTIIPRESG -KFLNOL 638
QY	890	NLP-EAMK	KDPGSESGRKAC	EVCGSHPOTQAL	EHOKTARCDPLFTVCBROGFLDRA 948
Db	639	HSPHEAMP	SFLGISTPMRKRI-----	LPRATL	DLMDPHN-----HFGIG-----QRRFVDSR 684
QY	949	TLKKHMLLA	HHQVPPRAPRH -P	ONIA	TLSLVPGSSSITPSGLSPFP 994
Db	685	STYALNMSRH	---PPROLGK	PSTSTSGATTEK	GPVAPRPIAPRP 728
RESULT 27					
09H8L4 PRELIMINARY; PRT; 498 AA.					
ID 09H8L4					

AC 09H8L4;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
CDNA FL113479 f1s, clone PLACE1003738, weakly similar to zinc finger protein 135.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
NCBI_TaxID=9606;
RN
RP
RT
TS
TI
TISSUE=PLACENTA;
RA Isoqei T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y., Nishikawa T., Nagai K., Sugano S., Takahashi-Fujii A., Hara H., Tanase T., Nomura Y., Togliya S., Komai F., Hara R., Takeuchi K., Arita M., Nabeckura T., Ishii S., Kawai Y., Saito K., Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y., Oshima A.;
RA "NEDO human cDNA sequencing project."
RT Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
EMBL: AK023541; BAB14602.1; -.
HSSP: P25490; 10BD.
DR InterPro: IPR000822; Znf_C2H2.
DR Pfam: PF00096; Zf-C2H2; 13.
DR PRINTS: PR00048; ZINC_FINGER.
DR ProDom: PD000003; Znf_C2H2; 2.
DR SMART: SM00355; Znf_C2H2; 13.
DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; 13.
DR PROSITE: PS00157; ZINC_FINGER_C2H2_2; 13.
DR DNA-binding; Nuclear protein; zinc-finger.
SQ
SEQUENCE 498 AA; 54872 MW; C704FCF55CCDAEIA CRC64;
Query Match 6.1%; Score 321.5; DB 4; Length 498;
Best Local Similarity 22.1%; Pred. No. 2,4e-11;
Matches 141; Conservative 70; Mismatches 237; Indels 191; Gaps 24;
371 KCRCAKVEGSSALQIHLRSHTEGRRPYCNVCNRPFTTGNLKVHFRHREKYPHYOWN 430
20 RCANCPKAYGALSKIKIHQNGHTGERPRACADCGKSPADS---VFKKHR--THAGIR 73
431 PNPVPHLDVYITSSGLRPYGMVPRPKAEAEAGTPGGGVKKRPLVAST--TALSTESLT 488
74 P-----YSGERGKAAVA-ELKDLRNHERSHT-----GERPFLSCSGKSFSSSSSLT 119
489 LLSGTSTAAVAPGLTFPNK--FVLMKAVERPSSKADENTPRSGSGAIVAGDSGATRMQL 547
120 CHORIHAAQKPYRCACGKGFOTLSSYOSHERTHSGEKPFLCPRCGRMFSPPSSFRHQR 179
548 SKLVTSLSWMLLTNHLKLTGSPRPYVLEPLG---ASPSSTKLQQLVERKIDRGAVAV 604
180 AH-----EGVK-----PYHCEKCGKDFROPADLA-MHRRVHTGDR----- 213
605 ASTASGAPTSAPAPSSSASGPNQVYICLVLSCPRALRLHYGNGHGERPFKCYGGRAF 664
214 -----PRKCLQCCOKTFVASWMLKRALVHSGGRPRFCSEGBGRAF 252
665 STRGNLRHNFVGHKTSPPARAQNSCPICQKKFTNAVTLQOVNRLHGGQ---IPNG---- 717
253 AERASLTGHSRVH-----SGERPFHCNACGKSFVYSSSLRKHERHRSSEAAGVPPAOELV 308
718 -GSALSEGGAQOENSSEGTASGSPFQPOSGQSPDEEMSEEEDEDEEDVTDID 776
309 VGLALPV--GVAGESSAAAPAGAGLDPPA----- 336
777 SLAGSGSBSGGEKALSVRODSEVSGAEAEVATSVAAPTTVKEMDSNEKARQHTLPRPPR 836
337 GLGLRPRESGG-----VMATQMVGVGM----- 358
837 PPDNIDHPQPMQGSIDVSGAMEEAKLEGISPPMALITQGEQSTPLVEELNLPEAK 896
359 ---TVEHNEJECODAGVREAPGPL-----EGAG-----EAGG 385

OY 897 KDPGSSGRKACVCGSPPTQTALEHOKTHPKDGLFTFCVFCRGLDRATLKKHML 956
DB 386 EADKKRPQVQRECKETFTSTMLLRHRSHPLELR-FPTQCGKSPSDRAGLRKHS-R 443
OY 957 AHHQVPPF-APHPQNIATLSLVPGCSSSLIPSPGLSPFP 994
DB 444 THSSVRYPTCPHCPKRAFLSASDLRKHERTHPVPMTPTP 482
RESULT 28
O9NS43
ID O9NS43 PRELIMINARY; PRT; 751 AA.
AC O9NS43;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Hypothetical zinc finger-like protein (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
NCBI_TaxID=9606;
RN
RP
RT
TS
TI
TISSUE=FROM N.A.
RA Kodoyanni V., Ge Y., Severin J., Krummel G.K., Grable L., Smith L.M.;
RA Kyvikstad E., Gordon L., Shannon M., Brower A., Olsen A.S.;
RT "Sequence analysis of a 1mb region in 19q13.2 containing a zinc finger gene cluster."
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
DR EMBL: AC074331; AAF88107.1; -.
DR HSSP: P08045; 1ZNF.
DR InterPro: IPR001909; KRAB.
DR Pfam: PF01352; KRAB; 1.
DR PRINTS: PR00096; Zf-C2H2; 18.
DR SMART: SM00349; KRAB; 1.
DR PROSITE: PS0805; KRAB; 1.
DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; 18.
DR PROSITE: PS00157; ZINC_FINGER_C2H2_2; 17.
KW DNA-binding; Metal-binding; Nuclear protein; zinc-finger.
FT NON-TER 1
SQ
SEQUENCE 751 AA; 86207 MW; 96583BA581C87FC CRC64;
Query Match 6.0%; Score 320.5; DB 4; Length 751;
Best Local Similarity 20.3%; Pred. No. 4,6e-11;
Matches 158; Conservative 71; Mismatches 287; Indels 263; Gaps 23;
OY 231 SSTKPLPLFSPRIKPAOTGKTATSSSSSSSSSSGAEPKQAFENLYNPLGSOHPFVGVG 290
DB 171 TDEP-----KPKGMEYGIITISGSKOKLPLGEKP-----HPCGCGSG 210
OY 291 RSHKPTPAPSPALPGSDIOLASPHLAFPPTGTLAOLGARGLEAAASBGLLPK-- 348
DB 211 PSY-----SPRLPLPN--VHTGECFQSGSHLRTQRIHNGEKILNRCHESGDCEFNKS 262
OY 349 -----NGSGELTGYGEVYISLEKRGKHKCFCAKVPSSDSALQIHLRSHTEGRRPKCNV 402
DB 263 FHSYOSNHTGEEKSY-----RCDSCGKGFSSSTGLIHYRHTGEKPKCEE 308
OY 403 CGNRPFTGNLKVHFNHREKYPHYVNMNPNVREHLDVITSSGLPYGMSV----- 453
DB 309 CGKCFSSQSNFQCHQVHTEKP-----YKCEGCGGFGSVNLRVQRYN 354
OY 454 ---PREKAEAEAGTPGGGVKKRPLVASTALSTATESLTLSTGTSTAAVAPGLTFNKFVL 510
DB 355 RGEKPYKCEE----- 377
OY 511 MKAV---EPKSKADENTPRSGSAIVAGDSGATRMQLSKLVTSLSWMLTNHLKST 567
DB 378 HORVHTGEKPYKCD-----YCGKGFSSHS-----PLCHRRVHT 411


```

Y      568 GSFE-----PVLTEPLGASSEFSKLDQLEYKIDROGA 601
b      412 GEKKYKCEACKCGFTRTNDLHIFRVHTGEKKYCKECCGFGSQASNLO--VHQNVHTGE 469
Y      602 VAVASTAGATTSAPAPSS-----SASGPNCVCYLRLVSCEPRALRLHYGGNGGERPK 656
Y      470 KRACETCGKGKFSSQSKTQIQRVTHTGKRYPCDYGCKDFSTSSNLKLQVHTHGKPYK 529
Y      657 CKVCGRAFSTRGNLRANFVGHKTSFAARAQNSCPTQCKKFTNAVTLQQHVKNHLAGQGIN 716
b      530 CEECGKGFSSWRSNLTAAHORVH---SGEKRPKCQCQDKSFSAIDFRVHQVRHTGEKPYK 585
Y      717 GGSALSBEGGAQAENSQSOSTASGPGSPFPQSOQPSPREEMSEEBEDEEDVDTDSD 776
b      586 CGVC--GKGSSQSSGLOSHORVHTGEKY-----RC 614
Y      777 SIAGRSGESGGEKAISVGDSEEVYSGAEEVATSYAAPTTVKEMDSNEKAPQHLLPPPP 836
b      615 DYCKGKGFYSOFLYHQGHGTGKEKKYCKECCGKGGRSL-----653
Y      837 PRDLMDHPQPEOGTSDVSGAMEERAKLEGISSPMALTOEGEGSTPLVEELNPENAK 896
b      654 ---MLRHQVRH--TGKPHICECGK-----AFSLPSNLR 684
Y      897 KDPESSGSRK--AEVCGSQSPFTQTALFEHOHTKHKDGFLPTVCVRGCFELDRATLK 953
b      685 VHGVHTREREKLFCKECCGKGFSSARLEAHORVHTGEKP-YKCDICDXKDFRRHSRLTYH 742

RESULT 29
57415 PRELIMINARY; PRT; 1615 AA.
D      057415 PRELIMINARY; PRT; 1615 AA.
C      057415;
T      01-JUN-1998 (TREMBLrel_06, Created)
T      01-JUN-1998 (TREMBLrel_06, Last sequence update)
T      01-JUN-2002 (TREMBLrel_21, Last annotation update)
T      Transcription factor RRB-1.
T      RRB-1.
S      Gallus gallus (Chicken).
S      Archaeoptera; Metacoa; Chordata; Craniata; Vertebrata; Euteleostomi;
S      Archosauria; Aves; Negunathae; Galliformes; Phasianidae; Phasianinae;
S      Gallus.
S      NCBI_TaxID=9031;
N      [1] |
N      SEQUENCE FROM N.A.
N      MEDLINE=98087435; PubMed=9427563;
X      Miyake J.H., Szeto D.P., Stumph W.E.;
P      "Analysis of the structure and expression of the chicken gene encoding
P      a homolog of the human RRB-1 transcription factor.";
P      Gene 202;177-186(1997).
P      -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
P      HEMBL; AF013754; AAB96584.1; -.
R      EMBL; P08045; 12NF.
R      InterPro: IPR000822; Znf_C2H2.
R      Pfam: PF000096; Zf_C2H2_15.
R      PRINTS; PR00048; ZINCINGER.
R      SMART; SM00355; Znf_C2H2_15.
R      PROSITE; PS00028; ZINC_FINGER_C2H2_1; 13.
R      PROSITE; PS01517; ZINC_FINGER_C2H2_2; 13.
R      DNA-binding; Metal-binding; Nuclear protein; zinc-finger.
R      SEQUENCE 1615 AA; 178279 MW; 5CA2F5FC0CPAlc79 CRC64;

Query Match          6.8%; Score 316; DB 13; Length 1615;
Best local Similarity 19.8%; Pred. No. 2,le-10;
Matches 234; Conservative 134; Mismatches 382; Indels 434; Gaps

134 GGGGLIASFKLGATLPPESTPAAPPPEPPPPGVGSHLNITPLEDELVALQOROIH 193
       |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
387 GSTGLLSLSPLEAATMGPFPSV-LPPIKE-----NIKL--SLQPFQKFI 430
       |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||

194 QNMNTEDI-----CRQVLLGSLGOTVGAPAPSELPGTCASSTKPLPLF 240
       |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
431 QPDSSIVYKRISNESALEADIOOLIKMAS-----SAPPOIS-LPPLSKASV-PVGSIF 483
       |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||

```

OY	241	S	-PLKPPQ--TGKTTTTSSSSSSSSSGABPPQKQAFHVLHVLGSGQHPFVGVGRSHK	295
Db	484	KHMPLEKPELPTPRIVATST	-----Pp-----PLISAQASGCI-----	520
OY	236	TPAPSPALPGSDOLI	-----ASPHLAPGTTGLLAOCIGAARGLEAASBGLKPK	348
Db	521	-----SESLPPPLRLIKNSVETSNSHLSQPAKSSPSQL	-----LQPYEP--LTQH	569
OY	349	NGSGELGYEBVLSL	-----EKPRG--HKCRPAK	377
Db	570	EKKTOLEODSITIALPLPLNMEAKIKQEVTEGDLKALITAGANKKAPMRYKVLPCRFQDQ	-----	629
OY	378	VEGSDSALQIHLKSHTGERPYKCNVC	-----GMR-FTTR-----GNL	413
Db	630	VFAFSQVLRAHRIHSLGISPYQCNICDYIAADKAAALRIHLRTHSGERPYYIKIGHPEYV	-----	689
OY	414	KVHFHRHKRYPHVQMNPRVPHLDYVITSSGLPYGMSVPPEKAEEDAG	-----	463
Db	690	KANCFERHLK-KRLKVTARDIEKIEVYNSMAEMWDFCSPDVCXLCGEDILKHYRLR	-----	748
OY	464	-----TPGGVEKRP	-----LVASTTALSA	483
Db	749	IHMRFHSGQKKRPFECKRCGTAFAKRNCRHHLKOLHVOEPEIENYILVDCSAQES	-----	808
OY	484	TESTLL	-----STGSTAABGLPTENKVL	510
Db	809	HTDAPLLEDSTYMDCKPIPTFLEPONGFSLGTSSSHVP	IKLEPMGNFMDEDEPLDFSQS	868
OY	511	-----MKAVEPSS	-----KADENPPGSEGSAIA	534
Db	869	KNLSAVQVQOENLIVSSPLSYFDCSMEPIDLSIPKYLKRNODIPGEARNQCELASVITDN	-----	928
OY	535	-----GVADSGSATBMQD	-----SKLYTSLPSMALLTNH--LKSTGSEFP	573
Db	929	AYNCOOCPPLGFANGSSEKNRAVGHOPQLGSKJLHLPVITISPLALGNSALLRPLRPPRP	-----	988
OY	574	YLLEPLGASPSSTSKIQOLVERKIDRGAAVAVASTAGAP	-----TTSAPABSSASGPNQC	629
Db	989	-----QPLRKPVPYTKELPPLA	-----SIAOIISSVSAPALLKTEADASPKAASSSTGC	1039
OY	630	-----VICLRVLSGPRAL	-----	646
Db	1040	DKSGNAKAKMTIVTAIQRDSNLPDQLQACDPERPDIADTGLRKRRKKGTNNKPKPLSS	-----RLHY	1099
OY	647	G--QHGERPPKCKVCVCAFSFTRGMLRAHFVGHKTSPPARA	-----QN-----	687
Db	1100	GVDLSSGEB	-----FASIEKMLTTTDNNKSPFLQSDTDFNEKESGONGTSEDK	1148
OY	688	-----	-----SCPICOKKFTAAVLOOHVHRMLGQ--JPNGSAL	721
Db	1149	ETPEBKLLRGKNTYISDCQKITTCPCRPVFSWASSIQIRMLTHTBSQADTEAPATGEV	-----	1208
OY	722	SEGGAAOENSSSEGSTASGPGSFPOQSOQSPSEEMSEEEDEEB	-----EED-----	771
Db	1209	LDITSCFEQEPVEVSELPSSECSPOBEOKADSPAEDEAEKADVEEGPEEDSVSNKL	-----	1268
OY	772	-----VTDEDSLAGRSGSESGEKAISVRGDSSEVSGAAEEVATSVAPATPYKE	-----	819
Db	1269	DLNTRASKLMDFKLASDOSAGSSSQTERRNHACDVCGRTEFAGALSRHKHAHIREDRKDE	-----	1328
OY	820	MDSNE--KAPQHTLPPPPPPPNLDHPQ--PME	-----QGSTDVGAAMEEAKLEGISPMA	872
Db	1329	RSSDESKSIODDAGAPSMQDSGLQBEESPMOLKVESPLDCATGKENESESIS	-----	1384
OY	873	ALTQBEESTIPVLVELNLPEAMKDPGESSG	-----RKACEVCGSGFPOTALEHQKT	927
Db	1385	-----EGEGETERKSTKSSSDQKTPKTDKASTAKADRKKVCYCNKRFWSLQDLTRHMS	-----	1440
OY	928	HPKQDPLFTVCFCROGFLDRATLKKNMLLHNOVPREAFRPN	-----	971
Db	1441	HTGERP-YKQTCERTFTLKLHSLVNHQRI-HQVKVATNRHNGKES	-----	1482

RESULT 30
 ID 99D2D7 PRELIMINARY; PRT; 1237 AA.
 AC 99D2D7;
 DT 01-JUN-2001 (TREMBLrel. 17, Created)
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE 4931408L03R1K protein.
 GN 4931408L03R1K
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 NC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NCBI_TaxId=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=TESTIS;
 RX MEDLINE=21085660; PubMed=11217851;
 RA Kawai J., Shingawa A., Shibata K., Yoshino H., Adachi T., Ishii Y.,
 Atakawa T., Hara A., Fukunishi Y., Konno H., Machi S., Fukuda S.,
 Aizawa K., Iawa M., Nishi K., Kiyosawa H., Kondo S., Yamana K. I.,
 Sakato T., Okazaki Y., Gojobori T., Bono H., Kusakawa T., Saito R.,
 Kadoya K., Matsuda H. A., Ashburner M., Batalov S., Casavant T.,
 Fleischmann W., Gaasterland T., Gissi C., King B., Koehle H.,
 Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
 Schiraldi L. M., Staudt F., Suzuki R., Tomita M., Wagner L., Washio T.,
 Sakai K., Okimoto T., Furuno M., Aono H., Baldarelli R., Bash G.,
 Blake J., Boftelli D., Bojunga N., Carninci P., de Bonaldo M. F.,
 Brownstein M. J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
 Guncicich S., Hill D., Hofmann M., Hume D. A., Kamuya M., Lee N. H.,
 Lyons P., Marchionni L., Mashima M., Mazzarelli J., Mombaerts P.,
 Nordone P., Ringwald M., Rodriguez I., Sakamoto N.,
 Sasakawa H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K. F.,
 Suzuki H., Toyooka K., Wang K. H., Welter C., Whitaker C., Wilming L.,
 Wysshaw-Borris A. O., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,
 Hayashizaki Y.;
 RA "Functional annotation of a full-length mouse cDNA collection."
 RL Molecule 409:685-690(2001).
 RL EMBL; AK019851; BAB31881.1; .
 DR HSSP; P17789; 2DRP.
 DR MGD; MGI:1925516; 4931408L03R1K.
 DR InterPro; IPR000345; Cyt_c_heme_bind.
 DR InterPro; IPR000822; znf_C2H2.
 DR Pfam; PF00096; zf-C2H2; 12.
 DR SMART; SM00355; ZNF_C2H2; 14.
 DR PROSITE; PS00190; CYTOCHROME_C; UNKNOWN_1.
 DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 8.
 DR PROSITE; PS0157; ZINC_FINGER_C2H2_2; 7.
 KW DNA-binding; Zinc-finger.
 SQ SEQUENCE 1237 AA; 130349 MW; 61016ARB31FE9D0 CRC64;
 Query Match 5.9%; Score 313.5; DB 11; Length 1237;
 Best Local Similarity 21.3%; Pred. No. 2.2e-10;
 Matches 257; Conservative 118; Mismatches 436; Indels 397; Gaps
 QY 13 GCGC-----EPAEKGGAASEHHPOVCAKCAQFSDPTFLAHONSCSTDPRVWIIIGQE 68
 DB 36 GGGGGGKRPESV--GGDS-----KREFAAAAEEN---DP-----E 65
 QY 69 NPSNSASSASAPREPGHRSQVM-----DTEHNSNPDDSGSS----- 103
 DB 66 SPAEASDHGHPQRPDPDTSIVANTVCPEQESLSLGDGSGEETKAGGIRKEPVGSGIM 125
 QY 104 ----GPPDPTWGP--RRGESSGQFLVAATGTAAAGGGGLIL-----ASPRLGATPLPPE 153
 DB 126 QNGCGGPEPELSLENPHSSAHASGN---AMKDKAVEGKTCILDFAHFGSEPGDHPDPPLPE 182
 QY 154 SPPAPP-----PPPPPP-----PPGVSGHNLPLLELRLVLDQQRD---HQMPPRQI 201
 DB 183 --PSQPRGGMAPPPESTPEFLAPENGSTLLP--PMSLLPQALQKQSCSPHNSGGLTOR- 238
 QY 202 CROVLLLSGATGATGAAASPS-----ELPCTGAASSTKPLPLFLFSDIKPAQOTGKT 251

Db	239	----	GPSSPBTACTIPASVSPPOVACVYSFKOSPCGHOSPASPAPYARSOCKPLKEEDBGV	293
QY	252	TAS----	SSSSSSSSGAEPKQAEFHLYHPGSHPSV-----	GGVGR-----SHK 294
Db	294	DKSPRSRQSPSSGAEMADEDESDNSPTSSSSSRPLKVRIKTITKSCGNTITVTVPSPSR	353	
QY	295	PTPAP-----	SPALGSGTDQLIASPHLA-----	FPCTGTGLA-----AQ 358
Db	354	DPPALAGAFLAETSPFLKLSPTPTPEGPVKVSVQLDGTRLKGTVLPVATIOMASTAM	413	
QY	329	CLGARGLEAAMSPG-----	-----	LL 345
Db	414	LMAASVARKAVLPBGNMNTSPKTKMSVLGVLPQLPKAEVTRGFSAGGKVGASVVMV	473	
QY	346	KPK-----	NGSGELGYEBVIS-----	-----SLEKPGR 359
Db	474	QPSKSGATPGTSGAG--GSVLSRTOSSLEAFENKILNSKLLPAYRPNLSPRAEGLALPRPG	532	
QY	370	HKCRFCAMVFGSDSLQILHRSHTGTERPKCVNCGNR--FTTRGNLKVHFHNRREK----	423	
Db	533	YRCLECGDAFSLKSLTARRIDRSMRIEVTCHNCARLVFPFNKCSILLNAREHKDGLVM	592	
QY	424	-YPHVQMPNHPVREHLDVYITSSGLPYGMSVPERKEAEEENATPBG-----	GYERKPLVA 476	
Db	593	QCSHLVMRVALDOMVGQPDITPLDR--VAVPR-----	YGPRLALPVLGEGEAVTS 642	
QY	477	STTALSTASLSLTL--LSTGSTAVAPRLPFENKVKLMAKAEPRKCAEDNTPRGSBSAIAG	535	
Db	643	STTTVAATEAPVLPPT-----	EPRAPPLASVYTCRCLCECKQCDRK-----	AG 687
QY	536	VA-----	DSSGATRMQLSKVLTSLPSMALLTNHL-----	KSTGSEPPRYVLEPIGAS---P 583
Db	688	MAHFGQGLRPPALGTSNVCSPSCP--MMLPNRCSFSAHQHTHKRNAPRVHPCEGGPNLOA	745	
QY	584	SETSKLOLVEKIDROGANAVALSTA----	SGAPTTSAAPRSSASGRNQCVIC--LRLVSCP 639	
Db	746	NQUTHIREACLFHSRRVGYRCPCAVFEGGVNSIKSHIQASHCEVTHKCPRLCPAERKSAP	805	
Db	806	SAHNLHYSQHPSLTQOAKLYKCAMCDVYFTHKPLLSHFDOH--LLPQVSVFKCPCSCP	864	
QY	694	KKFTNAVTLQOHR--MLGQILPNGSALSSEGGMAOENSEBOSTASGPFPOPOQOP	752	
Db	865	LTFQAKRTMLHKLKTHOSGRV--GEBAVKGAG-----	-----	CALLTPKTE-- 905
QY	753	SPPEEKSEEBEDEBEDVTDSDSLAGRSESGEKAISVRGDSEFVSGAEEEVATSA	812	
Db	906	PBE-----	-----LA-----	VSQAE-----A 916
QY	813	APTTVKEMDSNEKARQHTLPRPPRPDNDLHPORMEOGTSDVSGAMEBAKLEBISFMA	872	
Db	917	APATESSSSSEB-----	ELPSSPEPRPTPRARARGELGNGKING-----	----- 957
QY	873	ALTOBEGSTPLVBEHLNLEPA-----	MKKDPESSGRACCEVCGOSFOTALEEHOK	926
Db	958	---GGBPBGMTGCLCHSMCERODELYTHMKHEKMSVKKKPRCRLCEBSFCGAPSLRRVHR	1014	
QY	982	SSSIPSPG 989		
Db	1070	GGRAGPG 1077		
RESULT 31				
ID	096171	PRELIMINARY;	PRT;	734 AA.
AC	096171;			
DT	01-DEC-2001 (TREMblrel. 19, Created)			
DT	01-DEC-2001 (TREMblrel. 19, last sequence update)			

01-JUN-2002 (Tremblrel. 21, last annotation update)
 Unknown (protein for MGCL12654).
 Homo sapiens (Human).
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 NCBI_TaxID=9606;
 (1)
 SEQUENCE FROM N.A.
 TISSUE=UTERUS;
 Strausberg R.;
 Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
 EMBL; BC001777; AA07777.1;
 InterPro; IPR003309; Trp9_SCAN.
 InterPro; IPR000822; Znf_C2H2.
 Pfam; PF02023; SCAN; 1.
 Pfam; PF00096; Zf-C2H2; 13.
 Pfam; PF00003; Znf_C2H2; 6.
 Prosite; PS50804; SCAN_BOX; 1.
 Prosite; PS50028; ZINC_FINGER_C2H2_1; UNKNOWN_13.
 Prosite; PS50157; ZINC_FINGER_C2H2_2; 13.
 DNA-binding; zinc-finger.
 SEQUENCE 734 Aa; 82055 MM; B804766A0B008048 CRC64;
 Query Match 5.9%; Score 312; DB 4; Length 734;
 Best Local Similarity 19.4%; Pred. No. 1.4e-10;
 Matches 204; Conservative 75; Mismatches 329; Indels 444; Gaps 39;
 Y 11 LGPGCEPARERG-----DASEHHHPV-----CAKQ-----CAQFSD---PTFLAHQ 51
 b 6 LGSPDRAPPEDECPVAVKLEDSSEGEALMDPGPEARLARLRRCRYEATGPQELALQ 65
 Y 52 NSCCNDPPVYVILGGQENPSSASSAPREGRSGSRQVMDT-----EHSNPPD---SGSS 103
 b 66 RELC-----RQWLREVRKKEQMLELLVLEQFLGALPPEIQARQ 105
 Y 104 GPPDTWGERGESESSGOFLVATGTAAGG-----GGLIASPKLGA---PLPP 152
 b 106 G-----QRGSPREAAALVDGLRREPGRRRWVVOVGQEVLSKEMEPSSFPDPE 157
 Y 153 ESTAPPPPPPPPPPPVGSGLNPLILBELRYLQRO-----HQOMTQICQVLL 207
 b 158 TERPPRPGKPRP-----RTQESPLDQVKESEVED----- 192
 Y 208 LGSIGQVGAAPSSELPETGAASSTRKPLPLPSPIKAQTKTATSSSSSSSGAEP 267
 b 193 -----SDFESSGLPAATQESVPTLLP--EAAQCGVLDQIFPKSKGPBEP 237
 Y 268 KQAFPHLYPLGSPHPFSGVGRSHKPTPAPSPAL-----PGSTDQLIASPHLAP 319
 b 238 SWR-----EHPRALMHE-EAGGIF-----SPGFALQGISAGPGS-----VSPHLHVP 280
 Y 320 GTTGL--LAACIGAAKGLAAMAAPGLLKPKNGSGE-----LGVEVISLEKP 366
 b 281 WDLGMAGLSGQQSRSR--EGGFANHLILPSDLRSDDPDEDPCRGVGPALITTRWNSP 338
 Y 367 GGRH-----KCRCAKYFGSDSAQIQLHRSHTGERPKYKCNVGNRTTRGN 412
 b 339 RGRSGRGRSTGGGVVRRGCGDVGKVFQSRNLRHQKHITGERFVCSGGRSFRSSH 398
 Y 413 LKVFHHRREKYRHYOMNHPVREHLDVYITSSGLPYGMSVPRPKAEBAATPGGVERK 472
 b 399 LRLRLQHLHTERRPV-----KCRCAKYFGSDSAQIQLHRSHTGERPKYKCNVGNRTTRGN 420
 Y 473 PLVASTALSTATESLTLLSTGTSTAVAPGLPTENKPYLMKAVEPKSKADENTPPSESGA 532
 b 421 -----IVRSARLEHHRVHTGEOP----- 439
 Y 533 IAGVADGSATRMOLSKLVTSPLSPALLTNHLKSTGSPFPYVLEPLGASPSSTKIQOL 592
 b 440 -FRCAECGQSPFRSHN-----LQHORIHGDPPGP-----GAKP----- 472
 Y 593 VERKIDRGAVAVASTASGAPITTSAPAPSSA-----SGNQ--CVTCL 633

Db 473 -----PAPGAPPPPGPFPCSECRSARRAVLLEHQAIVHTCDKSGCYECG 519
 QY 634 RVLSCPRALRLHGOHGGERPFCKVCGRAPSTRGRLRAHPVGHKTSPPARAONSCPTCQ 693
 Db 520 ERFGRRSVLLQHRVRVHSGERPFACGCGSPFRQSRNLQHRRIH-----TGERPFACGEG 575
 QY 694 KKFETNAVTLQOHVYMHILGQGLPNGSGALSSEGGAAOENSSEOSTASGPGSPPOSOOPS 753
 Db 576 KAFRRQRTTLQHLKRVHTG-----EKP- 596
 QY 754 PEEMSEEEDEDEEDVDTEDESLAGSGESGKALSVRGDSSEVSGAEVATSVAA 813
 Db 597 -----FAC 599
 QY 814 PTTVKESDSNEKAPQHTLLPPPPPPNDLHPQMEQTSQSVSGAMEEAKLEGISSPMAA 873
 Db 600 PEGGQRFSGRLKTLRHQRTHTGKRP---YHGEGCGLGFTQVS----- 638
 QY 874 LQEGSGTSPFLVEELNLPKMKDPPGSSGRKACGCGSPFTQALTEHQTKTRKDP 933
 Db 639 -----RLTEHQRHTHTGERP--FACPEGGQSFROHANLTOHRRITHHTGERP 680
 QY 934 LEFTCFEGRQGLDRLATLKKHMLAHQVPPPA 965
 Db 681 -YACPEGCAFRQRRPRLTQH-LRTHREKPRFA 710

RESULT 32
 QYVQ09 PRELIMINARY; PRT; 744 AA.
 AC QYVQ09;
 DT 01-MAY-2000 (Tremblrel. 13, Created)
 DT 01-MAY-2000 (Tremblrel. 13, last sequence update)
 DT 01-JUN-2002 (Tremblrel. 21, last annotation update)
 DE Bowl protein.
 GN BOWL OR CG10021.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Ceiniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 Amandalides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
 George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 Brandon R.C., Rogers Y.-H.C., Blazek R.G., Champe M., Pfeiffer B.D.,
 Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 Abail J.F., Agbayani A., An H.-J., Andrews-Piannkock C., Baldwin D.,
 Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 Borkova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,
 Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 Dodson K., Dou P.L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
 Foster C., Gabrielian A.E., Gary N.S., Gelbart W.M., Glasser K.,
 Glodex A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 Hostin D., Houston K.A., Howland T.J., Wei M.-H., Idegawa C.,
 Jatali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclab J.M.,
 Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,

[illegible]

Db 1326 TMBEPATAPSTNSSLKTMIAQAAYGVGSLKEAVSSPFDESDQDLYPVAKVLVNAT--SQ 1384
 QY 846 PMEDGT---SDVSGAMEBEAKLEGISSPMAALTQEGESTPLVEINELPEAMKKDPE 901
 Db 1385 NMGNSTYFRPDSYDVAHNMHOSDE-EGIVASGASSESNNSGTEDYTSSSSSSPKKKSA 1443
 QY 902 SSGKCAEVCVGSPPOTVALEHOKTHPRDGLPTFCVFCQGLFDRATLKHKMLLHHQV 961
 Db 1444 APNRVSCPICORMFPPWSSSLRRHILITHGOKP-FKCSHCPLLTTKSNCDHLLRKGNV 1502
 RESULT 34
 061360
 ID 061360 PRELIMINARY; PRT: 962 AA.
 AC 061360;
 DT 01-AUG-1998 (TEMBLrel. 07, Created)
 DT 01-AUG-1998 (TEMBLrel. 07, Last sequence update)
 DT 01-JUN-2002 (TEMBLrel. 21, Last annotation update)
 DE CROL alpha.
 GN CROL OR CROL ALPHA OR CG14938.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CANTON S.
 RA D'Avino P.P., Thummel C.S.; EMBL/GenBank/DBJ databases.
 RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (by SIMILARITY).
 DR EMBL: AF020347; AAC15516.1; -.
 DR HSSP: P07248; IARD.
 DR FLYBASE: FBgn0020309; crol.
 DR InterPro: IPR000823; znf_C2H2.
 DR Pfam: PF000096; zf_C2H2; 18.
 DR PRINTS: PR00048; ZNCFINGER.
 DR ProDom: PD000003; znf_C2H2; 5.
 DR SMART: SM00355; znf_C2H2; 18.
 DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; 18.
 DR PROSITE: PS0157; ZINC_FINGER_C2H2_2; 18.
 DR DNA-binding: Metal-binding; Nuclear protein; Zinc-finger.
 KW SEQUENCE 962 AA; 106360 MW; 618DBCF99C9AAB CRC64;
 SQ
 Query Match 5.8%; Score 310; DB 5; Length 962;
 Best Local Similarity 20.1%; Pred. No. 2,6e-10;
 Matches 177; Conservative 83; Mismatches 321; Indels 300; Gaps 33;
 QY 161 PPPPPPPGSGGHLNPLILEELRYLQQRQIHOMQTEQICQVLLGSLGQTVGA--- 217
 Db 89 PPPPLP-----OVTSHAASAAAAAASTNMAVAA 121
 QY 218 -PASPSELPTGASSTRKPLPLFSPIKPAQTC---KTTASSSSSSSSSGAE----- 265
 Db 122 VMASANAASAAAAAASAGGLP-----PAISGNGGQGVTVITTSSTSSGSGTSGGTT 175
 QY 266 -----PRKQAFHLIYHPLGSOHPESVGVGRSHKPPAP--SPALPGS----- 306
 Db 176 TTAGELLMPMEGIIHGVDSGN-----GGNGGGQNALVADPGLTATGTHWCIDIGKM 229
 QY 307 ---TDLIASPHLAFPGTGLLAAQCLGARGLEAASPGLLKPKNGSGELGGEVI--- 360
 Db 230 FQFRYOLIV--HRYIYSEKRFKFCOVCG--QGFITSQD---LTRGKIHIG-GPMFTCI 280
 QY 361 -----SLEKRGGRHK-----CRFCAKVFGSDSALQIHLNRHTGEBRPKVCNCGN 405
 Db 281 VCFNVPANNSTLEENHKRHSHTDKPFACTICQKTFARKKELLDNHFRTSHGTETPRCQYCAK 340
 QY 406 RTTTRGNLKNHFNHKKRYR--VQMRPHRYENH--DIYITSSGLPYGMSYPRPEAAE 460
 Db 341 TTTKRHEHVMNHNKNGETPRHCDICKSKSTREKRYVNHVMHTGQTPHOCIDCGKK-- 397
 QY 461 EAGTPGGGVARKPLVASTTALSTATSESTLLSTGTSVAIVPGLPTFKFVLMKAVEPKSA 520

Db 398 -----YTRKEHLAN-----HMRSHT 412
QY 521 DENTPPSGSALAGVADSGATRMQLSKLVSLPSWALLTNHLKSTGSPFPVYLEBG 580
Db 413 NE-TPFRCB-ICCKSPSRKEH-----FTNH-----ILMHTG 441
QY 581 ASP-----SFT-SKLOOLVEKIDROGAVAVASTASGAPITTSAPASSSAGSPNCVCIL 633
Db 442 ETPHRCDFCSKTFTRKEHLNHN-RO-----HTGSPHNCSTCM 479
QY 634 RVLSCPRALRLHYGONGBERPCKVCGRAFTRGNLRAHFVGHKTSPPARAONSCPTCQ 693
Db 480 KTFPRKEHLNVNHRQHTGTEPRKCYCTKAFTRKD-----HMVNVNHRQHTGSPHNCSTCT 535
QY 694 KKFTNAYVTLQOHVNRHNLGQITPINGSALSSEGGAAQENSSEOSTASGSPSPPOQSDPS 753
Db 536 KTFPRKEHLNHNVRQHTG-----DSPHRCYSCKTFTRKEHLTNH 575
QY 754 PEEMSEEEDEDEDEEDVTDSDSLAGSGESGGEKALSVRGDSEEVGAEVATSVAA 813
Db 576 VRLATGDSPHKCEYQCKTTTRKEHL-----NNHNRQSSNDPHCCNVCK 620
QY 814 PTVKEMDSNEKARQHTLPPRPRLDHPORWEOGTSVSGAMEEBAKLEGISSPMAA 873
Db 621 PFTRKEHLINHMRSCHTGDPR----- 641
QY 874 LTQSGEGSTPLVEELNPEAMKKDPGESSGRKACEYCGSFPOTALEEHOKTHPKDGP 933
Db 642 FTCTCGSFPKGNLFLHQRSHTKGOEMERPRACEKCPKMFICKGHLVSHMRSGEKP 701
QY 934 LFTGCFRGFLDRATLKKHMLLHNHOV-----PPFAPHPQ 970
Db 702 -HACTLCSKAFVERGNLKRHKMKNHNDAMMPRPVPH-PQ 740

RESULT 35
ID 0960L6 PRELIMINARY: PRT: 736 AA.

AC 0960L6: 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE L046233p.
GN CG122299.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
NCBI_Taxid=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
RA Nunco J., Pacleb J., Paragas V., Park S., Phouanavong S., Wan K.,
RA Yu C., Lewis S.E., Rubin G.M., Ceiniker S.,
RA Submitted (Aug-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AY051994; AAK93418.1;
DR FLYBase: FBgn0032295; CG122299.
DR InterPro: IPR000822; znf.C2H2.
DR Pfam: PF00096; zf.C2H2_10.
DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; UNKNOWN_10.
DR PROSITE: PS50157; ZINC_FINGER_C2H2_2; 10.
DR DNA-binding; Zinc-finger.
SQ SEQUENCE 736 AA; 81970 MW; 0B1C198D547B4C2C CRC64;
Query Match 5.8%; Score 309; DB 5; Length 736;
Best Local Similarity 20.8%; Pred. No. 2.1e-10;
Matches 174; Conservative 91; Mismatches 283; Indels 288; Gaps 33;
QY 142 SPKLGATPLPPESTPAAPPPPPPPGVSGHLN-----PLIL-BELRVLQQRQIH 193

Db 17 TPILQOHTLLPQARBNPQOQRPQRPDLTFHOMCCAEFFVHNPLATLYQHMTLPHNERG 76
QY 194 QMO-----MTEQICROVLLIGSLQITGAPARSPSELPTGTGAASSKRPILP 238
Db 77 NQGOBOESPGESEDSYSWTFEPYC-----ELAEDSDSDGASGSDSSSSD--- 126
QY 239 LFSPIKPAQTGTATSSSSSSSSGAEPPKQAFPHLVNPLISQHPFS--VGVGCRSH--- 293
Db 127 --DDDDDDDDSSSSSSSSSSSSSVPTTSNSNTQOSQSVOPRHLVAGPGYNEPL 184
QY 294 -----KTPAPSPALPGSTDLIASPHLAFPTTGILLAAQCLGAAGLEA 339
Db 185 QMTDRESTSIFWQPTVAVTP-----LQOLLPRAPTVSPGLG--LQSTPIKRRGRRSN 237
QY 340 ASPLGLKPK-NG-----SGELGYGEVLSLEKPGGRKRCRCACAVFGSD 382
Db 238 IGAPWMDPALNGNOKCFQSTHCEASFPNAGDLSKVNRSHTNKP--FOCSICQTFTHI 294
QY 383 SALQIHLRSHTGRRPYKCNVCGNRFTRGMLKVHFRHREKYPH--YQMPHPVDEHLDY 440
Db 295 GSLNTHIRIHSGEKPYKCELCRKAFTQSSSLVMHMRSHSVKRPQCQCDKG----- 346
QY 441 VITSSGLPYGMSVPRPEKAEBAAGTPGGVERKPLVASTLTALSAATESLTLLSTGTAVAP 500
Db 347 FINYSSL-----LLHQKTHIAPTE-----TFICP 370
QY 501 GLPTNKFVLMKRAVERKSKADENTPPSGSALAGVADSGATRMQLSKLY-----TS 553
Db 371 BCE-----REFKAEALDEN-----MRNHTQELVYQCAICREA 403
QY 554 LPSWALLTNHLKS--TGSFPPVYLEPLGASPSETSILQOLVEKIDROGAVAASTASGAP 612
Db 404 FRASSELVQHMKNHMEKPTCSL-----CDR-----SFTQSG-- 436
QY 613 TTSAPAPSSSASGPNOCVICLRYLSCPRALRLHYGONGBERPCKVCGRAFTRGNLRA 672
Db 437 SLNIHMRHITGEPKFOCKLDCCTQASLSLVNMKIHAAGEKYPKPCICKSYSOQAYLTK 496
QY 673 HFVGKHTSPAARQNS-----CPICQKRTNNTVTLQOHV-MHL----- 710
Db 497 H1QAHOMASASASTSPGGLLVAKQPHETLYCTGCSLHADADALASHVHSQHNALLDTMK 556
QY 711 -----GGQIPNGSALSSEGG-----AAQENSSEOSTASGSPSPPOQSDPSPE 755
Db 557 QSGMNTAPGCAITDYKCSABEQAYVERVQCVLQAMNHOQOHQOQRPQOQOQHPOQ 616
QY 756 EBM-----SEEEB-----DEEBED-----V 772
Db 617 QOQNHQOQPHQWLPQPKLPAMDSTGEDEBERDEAEPRDEEBQDEBERPAEVKTEVL 676
QY 773 TDEDSLARGSESG--EKAISVRGD-----SEVSG--AEEVAT 809
Db 677 AAEDALTNPQYPIILGIEQIILDSDMYEDFGDMQVCGEEVFGGVVNEEEVYT 732

RESULT 36
ID 09VKF1 PRELIMINARY: PRT: 934 AA.

AC 09VKF1: 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE CROL protein
GN CROL OR CG14938.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
NCBI_Taxid=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;

[illegible]

QY	361	-----SSLEGGGRHK-----	CFECSKAFVSSDAIOJLHNSHTGERPKAWCGN	403
Db	281	VCNNVANNNTSLEHNMKRRHSIDKRFACI	IOCKTFARKHELNDHFNSTHETPRCOYCAK	340
QY	406	RETTGKMLKVHFNHRKRYR- --	VOHNRVREHL- -DYVITSSGLPYGMSVPRKAE	460
Db	341	TFPRKEHNMVHVNRKNTETPRHRCIDCK	SFTRKREHNVHNMHTGOTPRHOCVCGK- --	397
QY	461	EAGTPEGGVYRKPRVASTTALSNATESL	TLSTGTSTAVARGLPTFNKFLMKAVEKSKA	520
Db	398	-----TYRKEHLAN-----	-----HKRSIT	412
QY	521	DENTPRGSEGSALAGVADSGSATRMOLSK	IVTSLPRSMALLTNHLKSTGSPREPLYLEPLG	580
Db	413	NE-TPRPERCE- --ICGKSFSRKEH- --	-----FTNH-----	441
QY	581	ASP-----SET- SKIOOLYEKIDROGAV	ANVASTASGAPTTISAPARSSASGPNOCVTL	633
Db	442	ETPRHRCDFCSKTTRKEHLNHN- RQ	-----HTGSEPRHRCSTYCM	479
QY	634	RYLSCPRALYLNHGOHGERPRFKCVGGR	AFSTRGNLRAHFVGNKTPRAPAARONSCRPO	693
Db	480	KFTPRKEHLVNHTROHNGENPEKCTYT	CKATPRND- --HMVNHVROHNGESPRHKTYCT	555
QY	694	KKFTNAVTLDOAHVRYMLGGOTPRNGS	ALSBEGGAAOENSSSEOSTASGPGSPFOPOSQPS	753
Db	536	KFTPRKEHLTNHRLNLTG-----	-----	553
QY	754	PEHEMSEEEDEEEDVDYDEDSLARG	SESGEKAISVRDSEEVSGAEEVATSVAA	813
Db	554	-----DSPRKCYSOCKFTFRKEL- --	-----NNHMROHSDNDHNCVCWYCK	592
QY	814	PTTVKMDNSNEKARQHTLPPRRPR	NDLHPRQMEQSTDVSGAMEEAKLEGISSPMA	873
Db	593	PTTRKEHLNHNMSRCHGDRP-----	-----	613
QY	874	LYOEGEGTSPRYEELNLPRAMKKDP	GESGKACSEVCGSPRPTOTALJEHOKTHPKDR	933
Db	614	FTCECTGCKSPFLKGNLLFNORSHTK	GOEMERPRFCEKCRPKNETCKGHLVSHRSHSGEXR	673
QY	934	LEFTCVFCROGFIDRATLTKKMLNLAN	OV- --PPRARPRQ	970
Db	674	-HACTLCSKAFVERGNLKRMMKMNPR	ADMMRRPRVNR- PQ	712
RESULT 37				
Q9W4J1	ID	Q9W4J1	PRELIMINARY;	PRT; 1893 AA.
AC	Q9W4J1;			
DT	01-MAY-2000 (TREMBlrel. 13, Created)			
DT	01-MAY-2000 (TREMBlrel. 13, Last sequence update)			
DT	01-MAR-2002 (TREMBlrel. 20, Last annotation update)			
DE	PEB protein (GHI0905P).			
GN	PEB OR EG:66A1.1 OR CG12212.			
OS	Drosophila melanogaster (Fruit fly).			
OC	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;			
OC	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;			
OC	Ephydroidea; Drosophilidae; Drosophila.			
OX	NCBI_TaxID=7227;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=BERKELEY;			
RK	MEDLINE=20196006; PubMed=10731132;			
RA	Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,			
RA	Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,			
RA	George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,			
RA	Stinton G.C., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,			
RA	Burdton R.C., Rogers Y.-H.C., Blazey R.G., Champe M., Pfeiffer B.D.,			
RA	Man K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,			
RA	Abvill J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,			
RA	Ballev R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,			
RA	Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,			
RA	Botkova D., Botchan M.R., Bouch J., Brokstein P., Brotlier P.,			

1483 HCPPLFTTKSNCDRHLLRKHGAV 1505

RESULT 38

077275 PRELIMINARY; PRT: 1891 AA.

01-NOV-1998 (Tremblrel. 08, Created)

01-NOV-1998 (Tremblrel. 08, Last sequence update)

01-MAR-2002 (Tremblrel. 20, Last annotation update)

EG:66A1.1 protein.

PEB OR EG:66A1.1 OR CG12212.

Drosophila melanogaster (Fruit fly).

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

Ephydroidea; Drosophilidae; Drosophila.

NCBI_TaxID=7227;

SEQUENCE FROM N.A.

Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.

-1- SUBCELLULAR LOCATION: NUCLEAR (By similarity).

EMBL; AL031227; CAA20227.1; .

HSSP; P07248; 2ADR.

flyBase: FBgn003053; Pdb.

InterPro: IPR000822; Znf.C2H2.

Pfam: PF00096; Zf-C2H2; 12.

PRINTS; PR00040; ZINC2INGER.

SMART; SM00355; ZNF_C2H2; 13.

PROSITE; PS00028; ZINC_FINGER_C2H2_1; 11.

PROSITE; PS00157; ZINC_FINGER_C2H2_2; 10.

DNA-binding: Metal-binding; Nuclear protein; Zinc-finger.

SEQUENCE 1891 AA; 205368 MW; 9E882364C36BB9BF CAC64;

Query Match 5.88; Score 307.5; DB 5; Length 1891;

Best Local Similarity 19.0%; Pred. No. 8.3e-10;

Matches 265; Conservative 126; Mismatches 448; Indels 555; Gaps 59;

6 GSSSRGRCGRCRERGRGDASEEHHPVCAKCAQFSDPTFPLAH-----QNSC 54

226 GSSPQ--GQCLSGSGSGIGADGEMHKYLCRICEVVSATPHFTNHIRCHNYANGDPTENFT 283

55 C-----TDPYVAVILG----- 65

284 CRICRVLSSASSLIDRHYLVHTGERPNCRCYCHLTITTCNGMHRMHTKHOVAQOSQ 343

66 -----QGEMPS-----NSSASSAPP 81

344 SQQQQSLQQQQSQQQRRQQQQHQPQQQQQNPAAQQQLMGNTLSARASYESDASCSTVVS 403

82 EGRSRQVMDTEHSNPNPDSG-----SGPRPTWGERK----- 115

404 SGSHSRSSSSSLNNNNNSHRANNMLKLEEELEVESTEDODTEKQRKRTTINNNTISE 463

116 -----GEESSGQFLVATGTAAAGGGGLILASPKLGATPLRPSTPARPPPPPP-- 167

464 QOEDMDDEEDDADVAMLTSTR--DVATLLAGASASGA-----ASSPTTSPASPALI 515

168 ---PGVSGS-----HLNIPRLIEELNVL-----QQRQIHQOMQTEQICR-Q 204

516 LSCPAGASDFETLRALCVNLDAMHSDIPAKCRDCEYIFATNRQLQSH-----CRLRP 568

205 VLLLGSGQTVGAPASP--SELPGTGAASSTKPLDLFSFKPQOTGKTATSSSSSSSS 262

569 NMLAGGILPRLIGASSPLNNEEPDEDEHGDDEDL-----EQRERLASQSD- 614

263 GAERPRQAFNLYPLRIGSQHNF\$VGVGGRSHKPTP-----APSP 301

Db 615 -----FFHOLYLKIKRTTANGCAISHPSPRIKHEPADTKDLADISILNMTSSSS 663

QY 302 ALPGSTDLIASPH---LAPFGTGLLAQCLGAA-----RG-----LEAASPLK 346

Db 664 SFLRNFQSVNTPNSQYSLDGRDQEEEDADFTSEFRMKLRGPCKICTVFPMLRA 723

QY 347 PKNSGELGGEVYTSLEKDGRRKRCPCAKVFGSDSLQILHRSHTGERPYKCNVGNR 406

Db 724 LK-----GHNRYHLGAVGAGPRCANNCPYAVCDKALVHRMRTNGDRPYECAYCNYA 777

QY 407 FTTGNLKVHF-HRH-----REK-----YHVOAMP 431

Db 778 FTTKANCERHILNRHGKTSREEVKRALIYHPADAGCEDSKSRLGDLADTSRISPTP 837

QY 432 HPVP-----EHLDYVITSSG--LPYGSVP-----PEKAEENAGT 464

Db 838 PPPPVNESKSQLKHMMLGENHNLAFVNOQPPRLIQYKSLDQLVDKKPSAPAPQ0000QOE 897

QY 465 PGGG-----VERKPLVASTALATESLTLSTGTSTAVAPGLPTFNKFLKAV 514

Db 898 KSGALDESMVDLDLSKP-----TGASLITPAVPTTPAAVAP----- 935

QY 515 EPKSKADENTPPGSESGALGVADSGSATRMO-----LSKLVTSLSWA 558

Db 936 -----VTPGG-----VGTPLAAAIQQQLLAAQQLFGAGGEYMQQLFRSL----- 977

QY 559 LITNLKSTGSPF-FPVLEPLGASPSTSKLOQLYEKIDROGAVAVASTASCAPTISAP 617

Db 978 ---MFQSGTSGFPFPPMPAP--PPQANPEKPPVSPNKRINPMV--GVGVGVV----- 1026

QY 618 APSSASGPNOCYICLVLSCPRALRLHYGONGGRRPFCKYCGRAFSTGNLRAHF-- 674

Db 1027 ---PPGGVYKVIKKNVL-MPKQKQRYRT--ERPFCEHCSARFTLRSNMRHKKQ 1078

QY 675 -----VGKTSFARAQN----- 687

Db 1079 HPQFYAQRQSRGHHVMRGASNVAAAAAATAVMAGPGSGSGFNHGHGNG 1138

QY 688 -----SCRICQ-----KFTNAVYLQQHYR-----MILGQIPN--- 716

Db 1139 SHSGHGAIPSEVQKCALIAQOLKANKNTDLLQALAHGSSSVAGNPLHFGRPLNPS 1198

QY 717 ---GGSALSEGGAA-----QENSSEOST-ASGSPFPQPOSQOQSPSEEMS 759

Db 1199 MNGS--SGNGGATADDEDEPKLIIDEDENHDEVEADVDVF-----EEDDEEMD 1251

QY 760 EEEF-----DEEEEDVTDEDS-----LAGRSESGE-----KAISYRG 795

Db 1252 EPDEPELILIDQPAKEAEDEEDELPRLEOLGTREAAQKMAETILQALIKAGKPLSPRP 1311

QY 796 DSEEVGAEREVAT-----SYAAPT-----VKEMDSNE-KAPQHTL 831

Db 1312 TKENAPAMPVATTTMOEPALPATSTPNSSLKTMIAQAEVYKSLKLEVASFPEESODL 1371

QY 832 PPPPPPPNDHPPOPMQGT-----SDVSGAMEEAKLEGISSPMALTOGEGSTPLVE 887

Db 1372 VVYAKLYDNAT--SONMGFNSTFRSDVANHMQSDE-BELVAGSGASESNNSGTEEDTSS 1429

QY 888 ELNLPAMKKDPRESSGRKACVCGSPFTQTALEBHQTTHPRKGLPFTVCFRQGFIDR 947

Db 1430 SSSSEPKKSAYSILAPRNVSCPYCORMFPWSSSLRHHILTHGQKP-FKCSHCPLLETTK 1488

QY 948 ATLKKHMLLAHVOY 961

Db 1489 SNCDRHLLRKHGAV 1502

RESULT 39

096S24 PRELIMINARY; PRT: 725 AA.

096S24; 01-DEC-2001 (Tremblrel. 19, Created)

689 CPICOKFTNAVTLQOHVHRMHLGGQIPNGGSALSEGGAQENSSEOSTASGPGSFPQPQ 748

444 EFGALQDIKHQRFVHKLEK---RHKCSLCGKMFKKKSHVRNHLRTI

QY 406 RFTTRGNLKVHFHRHREKYPH-----VQMPHPV--PEHLDVITSSGL---- 447

Db 279 P F N S P A N L A R H R L T H T G E R P Y R C G D C G K A F T Q S S T L R Q H R L V H A Q H F P Y R C Q E C G V R F H R 338

